

MONITORAMENTO, INFRAESTRUTURAS E TECNOLOGIAS

novos desafios na
comunicação científica

Rogério Mugnaini
Fernando Modesto
Organizadores



Monitoramento, infraestruturas e tecnologias: novos desafios na comunicação científica

Rogério Mugnaini
Fernando Modesto
(orgs.)

SciELO Books / SciELO Livros / SciELO Libros

MUGNAINI, R., and MODESTO, F., eds. *Monitoramento, infraestruturas e tecnologias: novos desafios na comunicação científica* [online]. Salvador: EDUFBA, 2024. 340 p. ISBN: 978-65-5630-715-2. <https://doi.org/10.7476/9786556309156>.



All the contents of this work, except where otherwise noted, is licensed under a [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença [Creative Commons Atribuição 4.0](https://creativecommons.org/licenses/by/4.0/).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia [Creative Commons Reconocimiento 4.0](https://creativecommons.org/licenses/by/4.0/).

UNIVERSIDADE FEDERAL DA BAHIA

REITOR

Paulo Cesar Miguez de Oliveira

VICE-REITOR

Penildon Silva Filho



EDITORA DA UNIVERSIDADE FEDERAL DA BAHIA

DIRETORA

Susane Santos Barros

CONSELHO EDITORIAL

Titulares

Angelo Szaniecki Perret Serpa

Caiuby Alves da Costa

Cleise Furtado Mendes

Evelina de Carvalho Sá Hoisel

George Mascarenhas de Oliveira

Mônica de Oliveira Nunes de Torrénté

Mônica Neves Aguiar da Silva

Suplentes

José Amarante Santos Sobrinho

Lorene Pinto

Lúcia Matos

Lynn Alves

Paola Berenstein Jacques

Rafael Moreira Siqueira



PPGCI
PROGRAMA DE PÓS-GRADUAÇÃO
EM CIÊNCIA DA INFORMAÇÃO

ECQ
ESCOLA DE COMUNICAÇÕES E ARTES
UNIVERSIDADE DE SÃO PAULO

Rogério Mugnaini
Fernando Modesto

Organizadores

MONITORAMENTO, INFRAESTRUTURAS E TECNOLOGIAS

novos desafios na
comunicação científica

Salvador
Edufba
2024

2024, autores.

Direitos para esta edição cedidos à EDUFBA.

Feito o depósito legal.

Grafia atualizada conforme o Acordo Ortográfico da Língua Portuguesa de 1990, em vigor no Brasil desde 2009.

COORDENAÇÃO EDITORIAL

Cristovão Mascarenhas

CAPA E PROJETO GRÁFICO

Gabriela Nascimento

COORDENAÇÃO GRÁFICA

Edson Nascimento Sales,

REVISÃO

Mariana da Cruz dos Santos

COORDENAÇÃO DE PRODUÇÃO

Gabriela Nascimento

NORMALIZAÇÃO

Kátia Rodrigues de Oliveira

ASSISTENTE EDITORIAL

Bianca Rodrigues de Oliveira

EDITORIAÇÃO

Rodrigo Schlabitze e Cecylle Amaral

Sistema Universitário de Bibliotecas - UFBA

Monitoramento, infraestruturas e tecnologias : novos desafios na comunicação científica /

Rogério Mugnaini, Fernando Modesto, organizadores. – Salvador : EDUFBA, 2024.

9,5 MB [PDF].

Acesso em: <https://repositorio.ufba.br/handle/ri/41447>

ISBN: 978-65-5630-715-2

1. Comunicação na ciência. 2. Comunicação na tecnologia. 3. Pesquisa – Sistemas de Comunicação I. Mugnaini, Rogério. II. Modesto, Fernando.

CDU 001.8

Elaborada por Selma Matos

CRB-5: BA-001001

Editora filiada à:



EDUFBA

Rua Barão de Jeremoabo, s/n Campus de Ondina

Salvador - Bahia CEP 40170-115 Tel: +55 (71) 3283-6164

edufba.ufba.br

edufba@ufba.br

SUMÁRIO

Prefácio • 9

Apresentação • 15

Parte I - Monitoramento

Capítulo 1 Avaliação científica das Ciências Sociais e Humanas: caminhos possíveis a partir da literatura especializada • **21**

Henrique Baraldi Spera

Rogério Mugnaini

Capítulo 2 Desafios para o monitoramento do acesso aberto no contexto brasileiro • **43**

Ednéia Silva Santos Rocha

Fabiana Andrade Pereira

Capítulo 3 Explorando a relação entre *rankings* de universidades e o uso responsável de métricas e indicadores • **85**

Solange Maria dos Santos

Parte II - Infraestruturas

- Capítulo 4** BrCris: uma infraestrutura informacional aberta sobre os ecossistemas brasileiros de pesquisa científica e inovação • **129**

Vivian dos Santos Silva

Adilson Luiz Pinto

Washington Luis R. de Carvalho Segundo

Thiago Magela Rodrigues Dias

- Capítulo 5** Google Scholar como infraestrutura de avaliação científica: potencialidades e limites • **159**

Fabiana Andrade Pereira

Rogério Mugnaini

- Capítulo 6** Avaliação por pares em periódicos científicos: características e importância para evitar periódicos predatórios • **195**

Edna da Silva Angelo

Marlene Oliveira

- Capítulo 7** Repositórios institucionais de produção científica: agregando valor com a implementação de indicadores relacionais complexos • **223**

Rubenildo Oliveira da Costa

Lúcia Maria Sebastiana Verônica Costa Ramos

Parte III - Tecnologias

Capítulo 8 O registro de fontes de dados cientométricos • 243

Grischa Fraumann

Svantje Lilienthal

Christian Hauschke

Capítulo 9 Metadados na comunicação científica: mapeamento de modelos e aplicações • 265

Fernando Modesto

Amanda Ramalho

Capítulo 10 Análise das ferramentas de geração automática e semiautomática de metadados: possibilidades e limitações • 295

Jean Carlos Borges Brito

Dalton Lopes Martins

Capítulo 11 Comunicação científica no Brasil e a Plataforma Lattes • 321

Marcos Luiz Mucheroni

Valmir Batista Prestes e Souza

Sobre os autores • 343

PREFÁCIO

A publicação deste livro, sob a perspectiva do monitoramento, das infraestruturas e das tecnologias, é uma ótima notícia, já que analisa os novos desafios da comunicação científica. O livro chega no momento certo para ajudar a compreender estes tempos de mudanças transcendentais nos hábitos de pesquisa e na comunicação da ciência, nos quais uma ampla variedade de recursos de informação está disponível e acessível para a realização das atividades científicas. Um tempo em que se considera a inclusão de novos atores na geração de conhecimento, muitas vezes distantes do mundo da ciência, bem como o enorme desdobramento de meios e canais de comunicação, com o objetivo de tornar a disseminação dos resultados de pesquisa mais acessível, fluida, transparente e inclusiva.

No livro, a maioria desses aspectos é abordada em uma perspectiva rigorosa, analisando contribuições muito significativas. No primeiro deles, analisa-se o monitoramento das atividades de pesquisa, principalmente aquelas incluídas no âmbito das ciências sociais e das humanidades. Nesse sentido, é necessário destacar a importância de analisar essas disciplinas científicas, pois, tradicionalmente, foram menos analisadas sob perspectivas métricas, devido às singularidades de suas atividades de pesquisa, com padrões distantes das ciências naturais e tecnológicas, nas quais a pesquisa mais local e em línguas vernáculas não é comum.

A análise desses aspectos deve ser feita a partir dos Estudos Métricos da Informação (EMI), que fazem parte da ciência da informação, e que ao longo de sua trajetória passaram por diferentes etapas em seu desenvolvimento, tornando-se cada vez mais dependentes dos avanços e das metodologias de outras ciências e, de certa forma, das mudanças que ocorreram na sociedade como um todo. Essas sinergias resultaram não apenas na absorção de conhecimentos e metodologias provenientes de outras disciplinas científicas, mas também na incorporação direta de pesquisadores dessas disciplinas à atividade de pesquisa dos EMI. Esse grande caráter de interdisciplinaridade e transdisciplinaridade dos EMI foi fundamental para o seu desenvolvimento,

permitindo avanços espetaculares na abertura para novos objetivos de pesquisa, colocando os EMI no centro das atenções de atividades importantes para o desenvolvimento da ciência da informação, da política científica ou da avaliação da pesquisa, entre outros.

A chegada da ciência aberta, com a implementação de novos padrões de trabalho no mundo da pesquisa, está representando um desafio de grande escala para os EMI, que precisam se adaptar e fazer novas propostas aos desafios em que o mundo da ciência está imerso, de maneira muito mais diversa, transparente e inclusiva. Tanto a atividade de pesquisa quanto dos novos atores que estão começando a ter responsabilidades em todos os processos científicos, sem esquecer, evidentemente, os novos cenários em que a comunicação científica se desenvolve, e que cada vez têm menos a ver com os que foram habituais até agora.

As novas iniciativas que estão sendo realizadas a partir da Declaração de San Francisco sobre a Avaliação da Pesquisa¹, bem como os planos de ação e atividades que estão sendo realizados pelos grupos de trabalho dentro da Coalition for Advancing Research Assessment (CoARA)² e, recentemente, as propostas que surgiram a partir da Declaração de Barcelona³, representaram um enorme impulso para a ciência aberta na maioria dos campos. Esse impulso está sendo especialmente importante nos aspectos relacionados à avaliação científica, ao acesso livre e ao uso responsável das ferramentas provenientes dos EMI, principalmente os indicadores.

Tudo isso está obrigando os especialistas dos EMI a desenvolver seu trabalho dando espaço a novos recursos e fontes de informação, com o objetivo de monitorar a atividade científica sob perspectivas não consideradas até agora – como são as bases de dados abertas e gratuitas –, trabalhar com documentos multilíngues e ampliar a tipologia documental para todos os tipos de fontes, com o fim de elaborar produtos mais precisos, inclusivos e transparentes, que deem espaço à avaliação de iniciativas científicas mais locais, ou de países afastados dos incluídos na corrente principal. A discussão

1 Disponível em: <https://sfdora.org/read/>.

2 Conferir: <https://coara.eu/>.

3 Acessar: <https://zenodo.org/records/10958522>.

no livro dos *rankings* de universidades pode ser um bom exemplo para ilustrar este último comentário, pois, atualmente, a maioria deles se baseia em indicadores que não se ajustam a uma realidade científica tão diversa, não só nas atividades de pesquisa, mas também nas docentes ou de transferência de conhecimento, que são realizadas nos sistemas universitários de um grande número de países. Um exemplo disso é o que estamos desenvolvendo na Espanha para monitorar as atividades científicas do Sistema Universitário Espanhol, utilizando para isso mais de 40 indicadores e 8 dimensões, onde uma grande quantidade de dados provêm de fontes abertas⁴.

Dentre os aspectos que são desenvolvidos no livro, muitos estão relacionados às infraestruturas dentro dos processos de avaliação e monitoramento da atividade científica. Como mencionei no início, a ciência da informação – em geral – e os EMI – em particular – têm uma forte dependência das infraestruturas disponíveis, de tal maneira que os avanços e mudanças nelas serão de vital importância para o desenvolvimento dos EMI nas diferentes tarefas em que são empregados. Nesse sentido, o projeto Ecossistema de informação da Pesquisa Científica Brasileira (BrCris) é uma excelente iniciativa que está alinhada com as que comentei dentro da ciência aberta, e que pode permitir o monitoramento das atividades de pesquisa brasileiras de uma maneira muito mais precisa. A inclusão nesse monitoramento de todas as fontes de informação em acesso aberto – nacionais e internacionais – que incluem pesquisas das instituições brasileiras sob a perspectiva da pesquisa local – em língua portuguesa e em uma ampla tipologia documental, bem como o desenvolvimento de ferramentas de trabalho específicas para a construção e para o uso do sistema de informação – é o caminho para ter uma ideia completa e precisa da pesquisa que se realiza no Brasil e que permitirá complementar a que já foi obtida a partir da análise das fontes fechadas, principalmente, as bases de dados internacionais de índices de citações.

No livro e dentro da seção de infraestrutura, faz-se uma análise geral das infraestruturas digitais disponíveis para os estudos bibliométricos e cientiométricos, como as que geram – Clarivate Analytics, Elsevier ou Digital Science – e as suas bases de dados multidisciplinares – Web of Science,

4 Disponível em: www.iune.es.

Scopus ou Dimensions –, analisando as vantagens e desvantagens que cada uma delas representa na coleta e no fornecimento de dados para a obtenção de diferentes métricas. Como contraposição a essas fontes, explicam-se outras de características distintas, como é o Google Scholar, fazendo especial ênfase nas diferentes funcionalidades que oferece, bem como na grande variedade de fontes que recolhe. Evidentemente, a grande quantidade de informação que o Google gerencia faz com que as potencialidades desse recurso sejam imensas na hora de fazer estudos de produção científica exaustivos. No entanto, atualmente há certos problemas na precisão das buscas.

Outras das infraestruturas que inclui essa seção são os repositórios institucionais e o papel que desempenham na análise da produção científica das instituições às quais pertencem, já que incluem todos os produtos de sua atividade científica, embora, neste aspecto, deva-se destacar que em muitos casos esses produtos estão limitados pelos períodos de embargo impostos pelas editoras que não publicam em acesso livre. Ambas infraestruturas digitais, Google Scholar e os repositórios institucionais, com os novos padrões de comunicação dos resultados científicos que estão sendo impostos na ciência aberta, estão ganhando em reconhecimento e valorização para a análise da produção científica, pela sua maior capacidade para aportar dados de impacto e visibilidade de documentos escassamente considerados até agora nos processos de avaliação, como são os que fazem parte da literatura cinzenta, como os relatórios, as atas de congressos, as teses etc., bem como os provenientes de publicações periódicas locais de escassa visibilidade.

O papel da revista científica, e especialmente o das revistas predadoras, é outro dos aspectos que são abordados no livro. É importante que se faça, pois a proliferação dessas revistas, promovidas por editoras que utilizam práticas de legalidade duvidosa e escassa ética científica, têm uma grande demanda ante a imperiosa necessidade que os pesquisadores têm de publicar seus resultados científicos, com o objetivo de promover suas carreiras acadêmicas.

A terceira seção do livro concentra-se na análise das tecnologias e no papel importante que desempenham para tornar facilmente acessíveis os dados usados nos estudos de produção científica. A utilização dos enormes

recursos de informação disponíveis, com cada vez mais destes em acesso aberto, seria completamente impossível sem o desenvolvimento das tecnologias adequadas para sua gestão. Por isso, nessa seção, são abordadas a evolução de algumas dessas tecnologias e o seu impacto nos estudos bibliométricos e cienciométricos, como os desenvolvimentos das Interfaces de Programação de Aplicação (APIs), a geração automática e semiautomática de metadados ou as plataformas curriculares. A transição para a inteligência artificial e seu enorme impacto na realização dos EMI não é feita neste livro, mas deixa-se a porta aberta para que os autores utilizem essa nova tecnologia em uma futura publicação.

Finalmente, não quero deixar passar a oportunidade de dedicar algumas palavras de carinho e reconhecimento à dra. Dinah Aparecida de Mello Aguiar Población, professora da Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP), por seu grande trabalho no desenvolvimento da biblioteconomia brasileira e por suas relevantes contribuições à ciência da informação em nível internacional. Seus trabalhos em cientometria e produção científica na área da saúde foram pioneiros na implantação e no desenvolvimento dos EMI no Brasil. A partir de suas pesquisas e com o compromisso de outros pesquisadores da USP, em 1992 conseguiu criar o Núcleo de Pesquisa de Produção Científica. Atualmente denominado Grupo de Pesquisa em Métricas da Ciência e Tecnologia (CiMetrias) o grupo é especializado nos EMI e na análise dos novos fluxos da comunicação científica, sendo liderado pelos professores Rogério Mugnaini e José Fernando Modesto da Silva, com quem tenho o prazer de manter uma excelente colaboração há vários anos nos temas bibliométricos e cienciométricos, e que tem dado lugar a importantes resultados científicos.

Dr. Elias Sanz-Casado
Universidade Carlos III de Madri

APRESENTAÇÃO

O estudo da produção científica serve de insumo para diversas atividades relacionadas ao planejamento e à gestão do amplo sistema de ciência, tecnologia e inovação. Há mais de um século os documentalistas se pautavam em informações bibliográficas para compreender as práticas relacionadas à produção científica, enriquecendo assim o desenvolvimento de coleções de periódicos científicos. Não muito depois, as citações passam a compor o conjunto de informações úteis, capazes de evidenciar a intensidade do uso destes.

No bojo de tais ações, os catálogos passam a dar ao acervo maior amplitude e visibilidade, na medida em que não estão mais vinculados aos documentos físicos. Assim, a informação bibliográfica tem sua circulação e uso expandidos, fazendo com que a posse do documento deixe de ser o principal foco. O documento mantém sua importância, com base em atributos tradicionais que lhe creditam valor, como a temática, a autoria, o veículo de publicação e a indexação. Entretanto, as diversas transformações produzidas pela tecnologia de informação – o surgimento de fontes e sistemas de informação – sinalizam que, além de seu conteúdo, sua representação é crescentemente importante.

Os metadados, então, surgem como um elemento capaz de diversificar a descrição dos registros bibliográficos, multiplicando significativamente seu alcance, podendo ser vinculados a atributos diversos da autoria, assim como da instituição na qual foram produzidos, ou ainda da que os financiou. Outros atributos dizem respeito às fontes de informação em que se encontram, sendo assim, um elemento que permite a vinculação entre fontes, já que o documento passa a ser registrado em múltiplos sistemas.

Contudo, a identificação de documentos, ou de forma mais genérica e contemporânea, de objetos digitais, passa a contar com identificação única, trazendo confiabilidade na associação das informações dispersas, uma vez que os erros se multiplicam entre os usuários da informação. Os próprios autores requerem identificação única, assim como instituições, periódicos

científicos e agências financiadoras da pesquisa, pois precisam ter seu reconhecimento garantido.

Um complexo cenário informacional se consubstancia em uma infraestrutura descentralizada, porém altamente interconectada. O armazenamento e a recuperação da informação dependem de linguagens de marcação e codificação. Trafegar em ambientes digitais constituídos de mídias permite o rastreo de interações acadêmicas e sociais. A informação científica, então, deixa o ambiente estritamente acadêmico, misturando-se nos mais diversos contextos da sociedade.

O contexto altamente mutante em que são disponibilizadas as múltiplas informações resultantes da atividade de pesquisa científica permite que a produção científica tenha o seu devido reconhecimento, este resultante de seu processo de comunicação. A interação proporcionada nesse ambiente digital dinamiza a comunicação científica, assim como viabiliza a mensuração da circulação e o uso da informação.

A compreensão de tais aspectos deve ainda considerar o movimento de abertura da ciência, que perpassa todos os elementos mencionados, com destaque à tecnologia – com a abertura de *softwares*, *hardwares* e códigos – e ao acesso à literatura em acesso aberto, e, mais recentemente, aos dados de pesquisa. Ademais, deve-se considerar a integridade da pesquisa, que impõe o elemento da transparência no processo de geração do conhecimento. Se não bastasse, chama o cidadão para ser um ator participante do planejamento do projeto, justificando-a socialmente desde a concepção.

Assim, o monitoramento de processos avaliativos, do procedimento de abertura da ciência e dos parâmetros de uso responsável das métricas tornam-se ferramentas importantes. Lançando mão de inúmeras fontes de informação, passam a mensurar outros aspectos da vida acadêmica – por exemplo, aqueles registrados no currículo –, permitindo o delineamento do contexto em que se desenvolve o trabalho acadêmico, e que influencia a geração de conhecimento.

Em um âmbito mais amplo, a infraestrutura de informação abarca todos os recursos informacionais e tecnológicos supracitados, não se limitando ao território nacional ou institucional, a áreas do conhecimento, ou ainda, ao meio acadêmico. Apesar de possuírem importantes recursos – muitos

deles com longa tradição – e de pertencerem à iniciativa privada, iniciativas abertas de grande magnitude têm surgido. Os aspectos tecnológicos tornam-se, portanto, essenciais para garantia do acesso, da interoperabilidade, da gestão, do arquivamento, do compartilhamento e da mineração, que são viabilizados por sistemas diversos.

Nesse sentido, esta obra reúne questões críticas relacionadas aos novos desafios na comunicação científica, reunindo contribuições resultantes dos trabalhos realizados pelo Grupo de Pesquisa em Métricas da Ciência e Tecnologia (CiMetrias). Os textos não esgotam os temas tratados. Aliás, a intenção dos organizadores deste livro é a de constituir uma obra aberta que possa liberar uma contínua produção de saberes que possibilite novas explorações do tema.

As contribuições de convidados especialistas, de outras instituições e do exterior, permitiram o delineamento de um panorama mais amplo e mais palpável do objeto em questão. Os tópicos relacionados foram estabelecidos segundo os critérios discutidos com os autores e apoiados nos objetivos do grupo de pesquisa, fornecendo – dentro dos parâmetros estabelecidos – ampla liberdade de enfoque sobre o que consideram de interesse na abordagem do tema central.

Os organizadores

Parte I

Monitoramento

CAPÍTULO 1

Avaliação científica das Ciências Sociais e Humanas: caminhos possíveis a partir da literatura especializada

Henrique Baraldi Spera

Rogério Mugnaini

Avaliação qualitativa e quantitativa da ciência

Em uma conferência do British Council no verão de 1953, Karl Popper proferiu a seguinte frase a respeito do caráter científico de uma teoria: “pode-se dizer, resumidamente, que o critério que define o status científico de uma teoria é sua capacidade de ser refutada ou testada” (Popper, 1980, p. 5). O pensador austríaco havia estabelecido que era no momento da submissão à crítica, e não da verificação lógico-empírica, que uma teoria poderia ser considerada científica. Segundo esse princípio, que chamou de falseabilidade, se uma teoria puder ser refutada, então ela é científica. Por outro lado, se não oferece nenhuma possibilidade de crítica que leve à refutação por outros cientistas, então essa teoria não pertence à ciência.

Epistemologia não é o assunto que trataremos. Então, o que nos interessa nessa breve fala de Popper? É o que está subentendido. O pano de fundo. Afinal, a frase expõe um ponto fundamental: o ato de julgar o que é ou não “ciência” é uma prerrogativa dada somente aos próprios cientistas, aos “pares”, aos colegas que formam uma comunidade de “iguais”. Isto ocorre porque o fazer científico obedece a critérios formais muito específicos e rígidos, que exigem um nível de formação e de acúmulo de conhecimentos prévios tão vastos – teóricos, metodológicos, tecnológicos, terminológicos – que, na prática, a avaliação de um trabalho científico se torna uma tarefa de superespecialistas. Este é um dos motivos – embora não o único – que

faz com que o processo de avaliação da ciência como um todo forme uma estrutura interna tão perene e difícil de ser suplantada.

Historicamente, o estabelecimento de um pré-sistema de avaliação qualitativa da ciência teve início nas primeiras sociedades científicas da Europa da segunda metade do século XVII, como efeito do intercâmbio de textos entre os “homens de ciência” da época, principalmente na França e na Inglaterra, desencadeando um lento processo que levou à prática recorrente de averiguação mútua dos trabalhos originais, que recebiam comentários, apontamentos, críticas e correções – em outras palavras, aquilo que, com o tempo, convencionou-se chamar de revisão por pares (Zuckerman; Merton, 1971).

A revisão por pares desenvolveu-se nesses 250 anos, aprimorou-se muito, mas foi desde o início, e continua a ser, um sistema de autorregulação mantido por uma fechada comunidade de especialistas. Antes de ser mera etapa formal que delimita o período entre o final da pesquisa propriamente dita e a publicação dos resultados, a avaliação da produção científica pelos pares é parte constitutiva do próprio processo científico (Davyt; Velho, 2000). Nesse sentido, ela constitui, até hoje, uma prática tão imbricada no processo de avaliação que acabou por se estabelecer, de fato, como elemento constitutivo da ciência.

No entanto, a revisão por pares não é o único método de avaliação científica que conhecemos. Na segunda metade do século XX, outros procedimentos foram ganhando força, impulsionados por um formidável desenvolvimento científico e tecnológico e de um crescimento do volume informacional e documental sem precedentes na história. Isto abriu caminho para o desenvolvimento de outros mecanismos de mensuração que visassem compreender – e, se possível, controlar – a “dinâmica evolutiva” da ciência.

Áreas de estudos quantitativos da ciência despontaram das investigações sobre a evolução da atividade e da produção científicas a partir do pós-guerra, inaugurando a era da cientometria – o estudo dos aspectos mensuráveis de todos os processos envolvidos na cadeia de desenvolvimento científico. Sem dúvida, a seção da cientometria que mais prosperou foi a bibliometria, termo que deriva das palavras gregas “*biblion*” (livro) e “*metron*” (medida), para denotar a ênfase na mensuração dos aspectos relativos à

produção bibliográfica. Além disso, nenhum outro produto da bibliometria foi mais profícuo para a ciência da informação e para o desenvolvimento de métodos avaliativos da ciência do que as análises baseadas em contagem de citações, cujas contribuições práticas mais marcantes emergiram a partir de 1955, com os estudos de Eugene Garfield. É inegável o protagonismo do Science Citation Index (SCI) e do Fator de Impacto (FI) na proeminência da avaliação quantitativa da ciência, já que ambos eram vinculados ao Institute for Scientific Information (ISI), de Garfield.

A este respeito, Mugnaini (2016) recorda que a disponibilidade de dados de citação entre periódicos – como eram dispostos na publicação regular do Journal Citation Reports (JCR) desde a década de 1970 – viabilizou a realização de estudos da literatura, como aqueles realizados pelos documentalistas no início do mesmo século. Contudo, a utilidade de uma informação mais objetiva parece explicar a popularidade do FI – um indicador bibliométrico cujo propósito é revelar os periódicos científicos mais visíveis em um determinado campo da ciência. Para isto, o FI se baseia no número médio de citações dos artigos publicados durante um certo período. Em termos matemáticos, o FI de um periódico corresponde ao número de citações dos artigos nele publicados durante dois anos – citações recebidas no ano subsequente –, dividido pelo total de artigos publicados pelo periódico no mesmo período. Assim, os periódicos que possuem um FI maior são considerados mais “prestigiados” – isto é, supostamente detêm maior “influência” ou “impacto científico” naquele campo.

Importante frisar que, para figurar no cálculo do FI, os periódicos precisam estar indexados na base de dados Web of Science (WoS), atualmente propriedade da empresa Clarivate Analytics. Também é interessante notar que a reputação de indicadores de citação, como o FI ou, posteriormente, o Índice H, depende da avaliação qualitativa prévia a que os periódicos científicos são submetidos antes de ingressar em bases de dados como a WoS – afinal, antes da publicação, os textos passaram pela revisão por pares –, e que o uso desses indicadores seja percebido como um critério de avaliação mais “objetivo”.

A atribuição de qualidade a um artigo a partir do FI da revista gera grande discussão. Dentre outros motivos, por se tratar de uma avaliação

indireta, ou seja, em vez de analisar o artigo em si, pressupõe-se qualidades dele baseadas em bons atributos – igualmente pressupostos – da revista que aceitou publicá-lo. Parte-se do princípio de que, como o artigo foi aceito para publicação em uma revista de prestígio em determinada área, o próprio processo de seleção da revista avaliza a qualidade do artigo.

Por esse motivo, embora sejam considerados como um insumo importante no processo de avaliação, os indicadores de citação não substituem nem concorrem diretamente com a revisão por pares (The origin [...], 2024; Hicks *et al.*, 2015; Moed, 2005). A revisão por pares tampouco é imune a falhas e detém considerável cota de críticas¹, mas ainda é tida como o mecanismo mais criterioso de avaliação, porque somente nela os trabalhos submetidos são lidos e analisados, integral e individualmente, por cientistas gabaritados. Contudo, há que se considerar que a revisão por pares é um processo lento, caro e mais difícil de ser incluído em relatórios concisos. Nesse sentido, o uso de indicadores quantitativos possui atrativos e benefícios muito práticos. Alguns alegarão que os indicadores também conferem à avaliação uma almejada “neutralidade” – o que, diga-se, é uma afirmação igualmente controversa. Além disso – e neste caso parece haver consenso de ordem prática –, o uso de indicadores de avaliação simplifica e agiliza a análise de informações por parte das instâncias responsáveis por financiar projetos de pesquisa, destinar verbas para bolsas de estudo, conceder premiações etc.; em outras palavras, pelas atividades inerentes à implantação de políticas científicas, entendidas tanto em nível de grande amplitude (governamental) quanto de alcance muito focado (institucional). Neste aspecto, não há dúvida de que é mais fácil cotejar informações, estabelecer metas e padrões de produtividade, medir resultados, atribuir “valor” e tomar decisões “técnicas” a partir de números tabulados.

Avaliação da produção das Ciências Sociais e Humanas

No sistema avaliativo ora vigente, originalmente construído para atender às características das ciências naturais, a competição das Ciências Sociais e

1 Ver, principalmente, Chubin e Hackett (1990).

Humanas (CSH) por recursos simbólicos ou tangíveis, isto é, por prestígio ou por financiamentos, torna-se cada vez mais difícil. Uma evidência das adversidades pelas quais passam as CSH em comparação com as ciências duras é, por exemplo, a quase impossibilidade de demonstrar eficientemente o valor socioeconômico da pesquisa social aplicada, uma vez que seu impacto social nunca é facilmente detectável, quantificável ou mesmo estimável (Thelwall; Delgado, 2015).

O caso das CSH apresenta ainda aspectos editoriais, metodológicos, linguísticos etc. que interferem na aplicação dos indicadores de citação durante o processo de avaliação. Henk F. Moed, um dos principais pesquisadores dedicados ao estudo da análise de citação como instrumento de avaliação científica, reconhece em Derek de Solla Price o investigador que ressaltou que as “ciências” (naturais e da vida) e as “humanidades” – entenda-se, o conjunto das CSH – configuram dois domínios distintos de estudos, construindo conteúdos substancialmente desiguais, baseados em aparatos sociais diferentes de intercâmbio de informações, e evidenciando o modo como isto se reflete na literatura acadêmica das CSH (Moed, 2005). Nederhof (2006) salienta essas diferenças metodológicas expressas na própria trajetória de consolidação das CSH dentro do espectro das ciências, e destaca a heterogeneidade inerente ao conjunto de disciplinas de CSH, pelo fato de algumas se assemelharem às naturais e da vida, no que diz respeito às práticas de publicação e citação, mas outras não.

O estudo de Hicks (2004) alerta para as distorções observadas na avaliação bibliométrica das CSH, ao evidenciar “quatro literaturas das Ciências Sociais”, a saber: artigos de periódicos, livros – considerada a de mais alto impacto –, literatura nacional – relevante para impacto em nível regional – e literatura não acadêmica – importante para mensurar o conhecimento aplicado. Ocorre que o Social Science Citation Index (SSCI), indicador mais utilizado em avaliações, considera somente a primeira literatura, em formato de artigos, e ignora as demais formas de comunicação de resultados. Disto, permite-se então afirmar que o problema da avaliação das CSH passa de maneira decisiva pela questão das fontes de informação, ou, mais especificamente, às bases de dados que indexam literatura científica, como a Wof ou a Scopus, iniciando-se pelo fato de que o sistema de comunicação

científica baseado em artigos de periódicos não revela a verdadeira estrutura nuclear da produção das CSH, como tende a ocorrer nas ciências experimentais (Moed, 2005).

Em relação às fontes de informação científica, um ponto importante a ser reafirmado é que, tradicionalmente, as CSH ainda valorizaram a publicação de sua produção intelectual em formato de livro – ou capítulo de livro – em detrimento do formato de artigo de periódico (Benoit; Marsh, 2009; Hicks, 2004; Nederhof, 2006; Sivertsen; Larsen, 2012). Giménez-Toledo, Tejada-Artigas e Mañana-Rodríguez (2013) explicam que essa “preferência” decorre de necessidades concretas, pois o livro é essencial para as áreas das humanidades e algumas ciências sociais devido às características únicas da pesquisa nessas disciplinas. O livro oferece o espaço necessário para comunicar os resultados da pesquisa, acumular conhecimento, contextualizar o objeto de estudo, descrever variáveis e citar documentos de forma abrangente. Ao contrário de artigos ou trabalhos de eventos, os livros permitem uma abordagem mais detalhada, adequada às necessidades de pesquisa dessas áreas específicas.

Os autores destacam ainda que as características de disseminação das CSH não são consideradas pelos critérios comumente adotados pelos órgãos nacionais de avaliação científica, concluindo que os livros deveriam ser a “pedra angular dos processos de avaliação nessas áreas porque, caso contrário, [...] serão julgadas com base em hábitos de outras disciplinas e, portanto, estranhas à sua natureza” (Giménez-Toledo; Tejada-Artigas, Mañana-Rodríguez, 2013, p. 65-66, tradução nossa).

Se considerarmos que as principais fontes internacionais de informação bibliográfica científica – as bases de dados WoS e Scopus – indexam majoritariamente literatura veiculada em formato de artigo de periódico, e, logo, geram seus indicadores bibliométricos levando em conta artigos, podemos, desde logo, destacar a dificuldade que as CSH enfrentam nos processos de avaliação. Ora, a consequência natural dessa sub-representação em bases internacionais é que, por mais que publiquem (bons) livros, as CSH encontram fortes obstáculos para figurar de forma destacada nos relatórios de alto impacto baseados em indicadores de citação, porque livros “não contam muito”. Ademais, ao contrário do que ocorre nas ciências

naturais, as citações em CSH são mais dispersas (Hicks, 1999), ou seja, não estão concentradas em um núcleo circunscrito de revistas internacionais; não devemos esquecer que a noção de “núcleo” de publicações, desenvolvida por Bradford na década de 1930, constitui justamente a noção central que levou à construção do SCI e do FI (Sivertsen; Larsen, 2012), que valorizam sobremaneira os periódicos internacionais que habitam a “primeira zona de Bradford”.

Seguindo por esta direção, ao realizar um estudo bibliométrico sobre o impacto da pesquisa em ciência política na Irlanda, Benoit e Marsh (2009) igualmente constataram que as CSH possuem características e hábitos de publicação distintos, e salientaram que as análises de impacto em citações deveriam considerar também os livros como documentos citantes. A pesquisa apresentada pelos autores evidenciou que livros contribuem com uma parte importante da rede de citações que normalmente é limitada aos artigos indexados nas bases de dados.

A propósito dessas questões, Hicks (2004) afirma que avaliações de CSH amparadas em indicadores de citação disponíveis nas bases de dados internacionais tornam-se um desafio, porque os avaliadores têm que atuar sobre um conjunto de literatura “realmente confuso”, em que “[...] revistas em língua inglesa indexadas coexistem com a publicação de livros não indexados, a literatura nacional e a literatura não científica. [...] A literatura básica é menos claramente definida” (Hicks, 2004, p. 474, tradução nossa), o que nos leva, aliás, a outro ponto de debate que é a barreira linguística nos textos científicos.

Uma parcela significativa da produção científica em CSH se refere à análise de questões circunscritas às localidades em que os fenômenos sociais e humanos ocorrem, acarretando uma “orientação nacional e regional mais pronunciada” (Nederhof, 2006). Assim sendo, é comum que boa parte das pesquisas dessas áreas ao redor do mundo sejam publicadas em periódicos nacionais e redigidas em suas línguas locais (Hicks, 1999, 2004; Nederhof, 2006). Como as principais bases de dados internacionais indexam somente artigos escritos em língua inglesa – o idioma consensualmente adotado na difusão mundial da ciência, notadamente, as naturais –, muitos estudos sociais e humanos oriundos de países não anglófonos são automaticamente

excluídos dessas fontes de informação. Por isso, quase todas as disciplinas que compõem as CSH – economia e psicologia são exceções, segundo o estudo de Hicks (2004) – já largam em desvantagem em comparação a outras ciências avaliadas por indicadores bibliométricos de citação.

O estudo de Archambault e demais autores (2006) demonstra o “casamento” que pode ser observado entre a bibliometria e as bases de dados da então Clarivate e os desvios que essa vinculação acarreta em termos de avaliação. Estudando a cobertura da literatura em CSH nessas fontes de informação, os autores indicaram um viés de 20 a 25% a favor da produção científica em língua inglesa e que os periódicos franceses, alemães e espanhóis eram subestimados nas bases da empresa. A esse respeito, Hicks e demais autores (2015) destacam o caso de pesquisadores europeus da área de história que foram prejudicados em avaliações nacionais por não publicarem em periódicos indexados na WoS. Eles compunham o corpo docente de um departamento de Psicologia e insistiram que seus livros e a literatura em língua nacional deveriam ser igualmente considerados na avaliação.

Isto vem interferindo, inclusive, nos hábitos de publicação das CSH, cujos pesquisadores não anglófonos gradualmente estão passando a publicar mais em inglês e menos em sua língua nativa (Engels; Ossenblok; Spruyt, 2012).

Qualidade científica é um conceito multidimensional e não pode ser capturada unicamente pelo impacto de citações, e isto mais ainda se aplica às CSH.

Neste sentido é que Moed (2005) argumenta que indicadores de citação, por si só, não “avaliam” a qualidade da ciência, senão como instrumentos para prover “critérios de produtividade” que propiciam aos avaliadores “expressar suas visões de qualidade acadêmica mais explícita e claramente”. A principal função de um indicador direcionado às CSH, ainda conforme o autor, é “estimular o debate entre acadêmicos do campo que está sob investigação, a respeito da natureza da qualidade acadêmica, suas dimensões principais e suas operacionalizações” (Moed, 2005, p. 151, tradução nossa). Sivertsen e Larsen (2012) consideram que um indicador de citações aplicado às CSH possui alto potencial de uso, mas apesar da existência de algumas iniciativas neste sentido, um instrumento desse tipo ainda está por ser

criado. O grande desafio é que nas CSH “os padrões de publicação e citação não são tão concentrados em um núcleo de periódicos internacionais, como vemos nas ciências naturais” (Sivertsen; Larsen, 2012, p. 568).

Hug, Ochsner e Daniel (2013) também afirmam que o processo de avaliação das CSH ainda é uma questão em aberto, apesar de os recentes esforços em busca de adequação. Eles identificaram quatro pontos cruciais que resumem as apreensões dos cientistas sociais e que demandam atenção para o desenvolvimento de ferramentas eficazes de avaliação para essas áreas:

- crítica dos métodos empregados na avaliação da pesquisa, que são originários das ciências naturais;
- reservas quanto à quantificação da qualidade da pesquisa;
- temor dos efeitos da avaliação negativa dos indicadores de citação;
- falta de consenso sobre os critérios de qualidade entre diferentes disciplinas das ciências humanas e, conseqüentemente, a impossibilidade de comparação ou avaliação da qualidade da pesquisa.

Por conseguinte, os autores apontam também quatro pilares sobre os quais deveria sustentar-se uma estrutura de avaliação para as CSH:

- adoção de uma abordagem interna: o desenvolvimento de critérios e indicadores devem surgir nas próprias CSH – idealmente em cada disciplina ou subdisciplina;
- a qualidade da pesquisa não pode ser definida por medições simples: ou seja, indicadores devem estar vinculados a conceitos de qualidade. Todos os critérios de qualidade devem ser definidos por um ou mais aspectos – ou seja, conceituação da qualidade da pesquisa – e cada aspecto deve ser vinculado, se possível, a pelo menos um indicador quantitativo – operacionalização da qualidade da pesquisa. Essa abordagem esclarece o que está sendo medido por indicadores e permite a identificação de critérios de qualidade quantificáveis e não quantificáveis;
- criação de noções de qualidade explícitas: a ideia de qualidade vinculada a um indicador, uma ferramenta de avaliação ou um procedimento de

avaliação precisa ser feito de forma clara, indicando a direção para a qual a pesquisa deve evoluir;

- busca de consenso em relação aos critérios de qualidade: deve-se adotar uma abordagem que permita alcançar um consenso em uma disciplina/subdisciplina ou, pelo menos, determinar quais critérios são consensuais, e quais não são (Hug; Ochsner; Daniel, 2013, p. 370).

Esses preceitos traduzem bem os pontos essenciais levantados por Moed (2005, p. 151-152) para guiar o desenvolvimento de uma metodologia de criação de um indicador de desempenho para as CSH, que incluem:

- coletar os documentos que apresentam a visão de estudiosos do campo sobre como a avaliação de desempenho da pesquisa deve ser – e como *não* deve ser – realizada. Relatórios anteriores de comitês de revisão por pares são uma fonte inicial interessante para empreender este levantamento;
- envolver os pesquisadores das áreas em todas as etapas do processo de desenvolvimento de um indicador. Eles devem participar da criação de sistemas de classificação e de estruturação da própria produção de pesquisa;
- examinar empiricamente esta estruturação, do ponto de vista de sua consistência e do grau de consenso entre os acadêmicos. Questões relevantes que exigem mais esclarecimentos podem surgir desta etapa, frutificando estudos empíricos posteriores e mais detalhados, onde for o caso. Isto é essencial para conduzir a um “sistema de classificação” adequado das atividades do pesquisador e da produção científica, com vistas a estabelecer critérios comparativos corretos de desempenho acadêmico;
- o processo de desenvolvimento de um indicador para as CSH requer interatividade entre os especialistas em bibliometria e os acadêmicos da área em questão.

Como se pode notar, iniciativas de avaliação nas CSH são implementadas de maneira esparsa e adquirem particularidades específicas. Por essa razão, o esforço de monitoramento de propostas ou experimentos é necessário, analisando políticas ou revisando a literatura.

A avaliação das CSH retratada na literatura

Com o propósito de examinar a literatura especializada para identificar o desenvolvimento e o tratamento dispensado à avaliação científica das CSH, realizamos uma revisão de literatura, a partir de um método de levantamento e seleção bastante específicos. As fontes de informação utilizadas foram a base principal da WoS e a rede digital ResearchGate (RG). No caso da WoS, a justificativa foi o fato de esta ser uma das fontes mais utilizadas em sistemas de avaliação pelo mundo. Já no caso do RG, o intuito foi incluir uma rede digital aberta de cientistas como fonte alternativa de recuperação de literatura, partindo de uma consideração da Lei de Dispersão de Bradford: diante da necessidade de obter informações, a maioria dos pesquisadores efetua buscas em poucas fontes, sendo que os padrões de dispersão de informação podem ser afetados pelo comportamento de busca dos cientistas (Bigdeli *et al.*, 2013). Partindo deste princípio, era importante expandir para uma fonte alternativa de literatura, na qual seria possível estabelecer conexões temáticas entre autores, mesmo que supostas a princípio. A ideia central foi comparar os levantamentos do RG e da WoS, ou seja, uma rede de autores e uma busca livre em uma fonte bibliográfica consagrada e fartamente estudada, fazendo emergir daí novos dados relevantes para análise. Partimos de um autor específico no RG, com produção consistente no tema da avaliação das CSH, e expandimos o conjunto de autores identificando suas conexões – autores que o seguem e são seguidos por ele na rede. Desse grupo de autores, extraímos a produção bibliográfica recuperada no próprio RG e a comparamos com a indexada na WoS.

Realizados os processos de levantamento e seleção da literatura recuperada por critérios puramente bibliométricos nessas fontes, trabalhamos com um *corpus* de análise composto por 107 artigos sobre avaliação das CSH², os quais passaram por uma segunda etapa de seleção, agora de cunho qualitativo. Revisamos individualmente a temática e a proposta metodológica (teórica ou empírica) dos artigos do *corpus* e, com isso, chegamos a um conjunto de dez artigos (Quadro 1) que apresentavam, ao mesmo tempo,

2 Para o leitor que desejar conhecer em detalhes os passos metodológicos para seleção de autores, levantamento da produção e seleção do *corpus* de análise, ver: Spera (2017).

características que chamamos de “reflexivas” – presença de argumentação crítica e visão questionadora sobre a avaliação das CSH – e “propositivas” – apresentação original e detalhada de estudo empírico e/ou proposta prática de avaliação das CSH.

Quadro 1 - Artigos do *corpus* com características reflexivas e propositivas

Referências dos artigos analisados
DONOVAN, C.; BUTLER, L. Testing novel quantitative indicators of research ‘quality’, esteem and ‘user engagement’: an economics pilot study. <i>Research Evaluation</i> , Oxford, v. 16, n. 4, p. 231-242, Dec. 2007.
GIMENEZ-TOLEDO, E.; ROMAN-ROMAN, A. Assessment of humanities and social sciences monographs through their publishers: a review and a study towards a model of evaluation. <i>Research Evaluation</i> , Oxford, v. 18, n. 3, p. 201-213, Sept. 2009.
GIMENEZ-TOLEDO, E.; ROMAN-ROMAN, A.; ALCAIN-PARTEARROYO, M. D. From experimentation to coordination in the evaluation of Spanish scientific journals in the humanities and social sciences. <i>Research Evaluation</i> , Oxford, v. 16, n. 2, p. 137-148, June 2007.
HUG, S. E.; OCHSNER, M.; DANIEL, H.-D. Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history. <i>Research Evaluation</i> , Oxford, v. 22, n. 5, p. 369-383, Dec. 2013.
NEDERHOF, A. J. Bibliometric monitoring of research performance in the Social Sciences and the Humanities: a review. <i>Scientometrics</i> , [Budapest], v. 66, n. 1, p. 81-100, 2006.
OCHSNER, M.; HUG, S. E.; DANIEL, H.-D. Setting the stage for the assessment of research quality in the humanities. Consolidating the results of four empirical studies. <i>Zeitschrift Fur Erziehungswissenschaft</i> , [s. l.], v. 17, p. 111-132, 2014.
OCHSNER, M.; HUG, S. E.; DANIEL, H.-D. Four types of research in the humanities: setting the stage for research quality criteria in the humanities. <i>Research Evaluation</i> , Oxford, v. 22, n. 2, p. 79-92, June 2013.
THELWALL, M.; DELGADO, M. M. Arts and humanities research evaluation: no metrics please, just data. <i>Journal of Documentation</i> , London, v. 71, n. 4, p. 817-833, 2015.
ZUCCALA, A. Quality and influence in literary work: evaluating the ‘educated imagination’. <i>Research Evaluation</i> , Oxford, v. 21, n. 3, p. 229-241, Sept. 2012.
ZUCCALA, A. et al. Altmetrics for the humanities: comparing Goodreads reader ratings with citations to history books. <i>Aslib Journal of Information Management</i> , [s. l.], v. 67, n. 3, p. 320-336, 2015.

Fonte: elaborado pelos autores.

É interessante de início perceber que sete dentre as dez pesquisas reforçam um enfoque misto acerca dos métodos de avaliação, isto é, combinam em suas propostas tanto métodos quantitativos, quanto qualitativos. Somente três trabalhos denotam uma prevalência entre método – seja quali ou quanti – em suas análises. Da mesma forma, oito dos dez artigos são

fruto de pesquisa empírica, evidenciando forte tendência dos estudos aplicados na literatura sobre o tema.

Os Quadros 2 e 3 diferenciam dois conjuntos de artigos: o primeiro reúne sete artigos que apresentam proposições práticas de avaliação, identificando meios e instrumentos para a implementação do processo; enquanto o segundo consiste em três artigos de um mesmo grupo de autores, que identificam conceitos de qualidade mais adequados para as áreas de CSH.

Do Quadro 2, podemos apontar o estudo de Donovan e Butler (2007), em que um levantamento de dados qualitativos – “reputação” (*esteem*) e “envolvimento” (*user engagement*) – foi proposto a especialistas de cinco departamentos de economia na Austrália como medida de avaliação da qualidade da pesquisa – dados de *user engagement*, aliás, foram encorajados, mas os de *esteem* foram rejeitados pela comunidade científica, como não tão dignos de crédito para este fim. As mesmas autoras apresentam também propostas de análises bibliométricas, em que ressaltam as limitações da avaliação quantitativa das CSH, propondo “opções” de adequação das análises de citação, em consonância com outras proposições (Nederhof, 2006; Zuccala, 2012; Zuccala *et al.*, 2015).

Thelwall e Delgado (2015) desencorajam o uso de indicadores “padronizados” de citação e propõem que os próprios pesquisadores elejam criativamente os dados que são reveladores do valor científico das pesquisas. Por fim, o artigo de Zuccala (2012), que, assim como os anteriores, não resulta da aplicação de projeto de avaliação, defende proposta baseada em combinação de dados qualitativos (revisão por pares) e quantitativos (indicadores de citação), porquanto medem dimensões diferentes: a revisão por pares desempenha papel crítico na percepção da qualidade da literatura científica em CSH, enquanto as citações de livros e artigos de periódicos podem ser usadas para rastrear padrões de influência.

Já Zuccala e demais autores (2015) propõem uma avaliação por dados quantitativos, extraídos das classificações dos leitores do Goodreads, no intuito de medir o impacto mais amplo de livros acadêmicos publicados na área de história, revelando o potencial de uso de indicadores altmétricos para identificar o impacto social de livros de CSH fora da academia.

O artigo de Gimenez-Toledo e Roman-Roman (2009) revisou projetos desenvolvidos em instituições de países que haviam aplicado metodologias e indicadores diversos para avaliar livros, combinando, para isto, dados quantitativos e qualitativos – FI ou outros indicadores de citação, revisão por pares e outros mecanismos de seleção de manuscritos, conformidade com padrões de publicação internacionais, divulgação da pesquisa em catálogos e bancos de dados; Gimenez-Toledo, Roman-Roman e Alcain-Partearroyo (2007) igualmente propõem um modelo de avaliação de periódicos de CSH combinando dados qualitativos e quantitativos, com base em diversos indicadores de qualidade e prestígio editorial, visibilidade internacional, revisão de pares e impacto.

Quadro 2 - Tipo de proposição prática de avaliação e formas de implementação

Tipo de proposição	Meio/instrumento de implementação	Nederhof (2006)	Donovan e Butler (2007)	Gimenez-Toledo <i>et al.</i> (2007)	Gimenez-Toledo e Roman-Roman (2009)	Thelwall e Delgado (2015)	Zuccala (2012)	Zuccala <i>et al.</i> (2015)
Adaptação da análise de citação	Citações a documentos fora da WoS (ou todas as fontes constantes das referências)							
	Peso para periódicos ou editores							
	Aumento da janela de citações							
	Documentos cocitados para identificação da influência e especialização de autores							
Avaliação de periódicos	Fator de Impacto							
	Indexação							
	Peer review, conteúdo do texto							

Tipo de proposição	Meio/instrumento de implementação	Nederhof (2006)	Donovan e Butler (2007)	Gimenez-Toledo <i>et al.</i> (2007)	Gimenez-Toledo e Roman-Roman (2009)	Thelwall e Delgado (2015)	Zuccala (2012)	Zuccala <i>et al.</i> (2015)
Avaliação de monografias	Prestígio da editora							
	Informações sobre a editora (do WorldCat.org)							
	Peer review, conteúdo do texto							
	Indexação em bases de dados							
	Presença em repositórios de instituições bem ranqueadas							
	Resenha (mensurar o percentual de sentenças positivas)							
	Classificações dos leitores do Goodreads							
	Duas formas de avaliação do impacto sociométrico: com “libcitação” (benefício “cultural” de um livro) e com “altmetria” (absorção cultural através das mídias sociais)							
Reconhecimento e impacto	Resultados acadêmicos: refletidos em outra pesquisa, dentro ou fora da disciplina							
	Resultados mais amplos: refletidos fora da comunidade acadêmica							
	Impacto acadêmico: impacto em pesquisa, dentro ou fora da disciplina							
	Impacto mais amplo: impacto fora da comunidade acadêmica							
	Reputação do pesquisador: premiações, honrarias, eleição para sociedades científicas, participação em conferências, atividades em revistas da área (editor científico, comitê científico, parecerista), envolvimento como pesquisador visitante em instituições estrangeiras (ou recebendo esses pesquisadores em sua instituição)							
	Envolvimento na pesquisa por órgãos governamentais: projetos de pesquisa, consultorias, relatórios e pareceres técnicos							

Fonte: elaborado pelos autores.

Por sua vez, o Quadro 3 traz as contribuições de três autores, cujos artigos destacam conceitos de qualidade para avaliação da pesquisa nas humanidades. Hug, Ochsner e Daniel (2013) utilizaram o método qualitativo Delphi, que envolve sucessivas rodadas de discussão entre especialistas, como ponto

de partida para identificar critérios consensuais de qualidade entre cientistas de três disciplinas de CSH – que culminou com a identificação de dois critérios “não consensuais” em todas as disciplinas, a saber: “produtividade” e “relação com impacto na sociedade”, evidenciando incompatibilidade de critérios entre avaliadores e acadêmicos. Outros dois artigos de Ochsner, Hug e Daniel (2013, 2014) resultaram, como continuidade dos estudos iniciais, de um projeto aplicado em universidades suíças, onde se desenvolveu uma proposta de avaliação de pesquisa em CSH também concentrada na questão da consensualidade, e cujos resultados indicam proeminência de uma avaliação de pesquisa por meio de critérios qualitativos - concepções de pesquisa e conceitos subjetivos de qualidade previamente coletados junto a especialistas por meio de entrevista –, sendo que indicadores ligados a essas noções de qualidade são considerados limitados, mas podem ser usados para apoiar o processo de revisão por pares.

Quadro 3 - Conceitos de qualidade para avaliação da pesquisa nas humanidades

Hug; Ochsner; Daniel (2013)
(1) Abordagem de dentro para fora. O desenvolvimento de critérios e indicadores deve emergir das próprias CSH, em cada disciplina ou subdisciplina, para que critérios únicos de qualidade possam emergir; (2) abordagem coerente e prática. A qualidade da pesquisa não é definida apenas por suas métricas: os indicadores devem estar relacionados a conceitos de qualidade. Todo critério de qualidade deve ser definido e especificado por um ou mais aspectos (conceituação da qualidade da pesquisa) e cada aspecto deve estar ligado, se possível, a pelo menos um indicador quantitativo (operacionalização da qualidade da pesquisa). Tal abordagem esclarece o que está sendo medido por indicadores e permite a identificação de critérios de qualidade quantificáveis e não quantificáveis; (3) conceitos explícitos de qualidade. As noções de qualidade que fundamentam um indicador, uma ferramenta de avaliação ou um procedimento de avaliação devem ser formuladas de maneira explícita. Ao desenvolver indicadores para medição de qualidade, deve haver entendimento geral entre os pesquisadores sobre o que é uma boa pesquisa; (4) esforço por consenso. Para superar a falta de concordância em relação aos critérios de qualidade, deve-se adotar uma abordagem que permita chegar a um consenso em uma disciplina/subdisciplina, ou que permita pelo menos detectar os critérios consensuais e aqueles que não são.
Ochsner; Hug; Daniel (2013)
Crítérios de qualidade da pesquisa: (1) continuidade; (2) inovação e originalidade; (3) rigor; (4) reflexão e crítica; (5) intercâmbio científico; (6) inspiração; (7) conexão com a sociedade; (8) diversidade e variedade; (9) atualidade; (10) abertura e integração; (11) autonomia; (12) bolsa de estudos; (13) conexão entre ensino e pesquisa; (14) motivação intrínseca; (15) produtividade.

Ochsner; Hug; Daniel (2014)

(1) uma vasta gama de critérios de qualidade deve ser aplicada para avaliar adequadamente a qualidade da pesquisa em ciências humanas; (2) há critérios compartilhados que são consensuais em todas as disciplinas estudadas; (3) as disciplinas não devem ser agrupadas, pois encontramos critérios disciplinares específicos; (4) com um certo cuidado, indicadores de pesquisa ligados aos critérios relevantes podem ser usados para apoiar os especialistas em avaliações de pesquisa (revisão por pares informados).

Fonte: elaborado pelos autores.

Considerações finais

A revisão qualitativa do *corpus* faz erigir um componente individual de percepção que difere da interpretação embasada nas evidências bibliométricas.

Além da contagem de características formais dos documentos – que sem dúvida ajudam o leitor a revelar aspectos de grande relevância –, a leitura proporciona uma interpretação de outra ordem: a imersão na linguagem do autor, o contato com a comunicação verbal explicitada no texto, mas, igualmente, os significados não evidentes, as mensagens que não são claramente enunciadas – absolutamente incompreensíveis pelos instrumentos estatísticos. Neste sentido, o exame dos artigos selecionados, como conjunto, sugere que a comunidade de cientistas sociais que se dedicam ao problema da avaliação se encontra em um ponto de sedimentação de “massa crítica”. Vislumbra-se, assim, que a literatura esteja em sentido de expansão e que exista uma inclinação latente em se produzir cada vez mais estudos “de resultado”, isto é, de análise de novas aplicações de avaliação das CSH, baseada em indicadores bibliométricos e qualitativos. Essa conjectura poderá ser verificada, por exemplo, na possibilidade de se expandir a coleta de informações bibliométricas deste *corpus*, agregando-se à produção consolidada de 2016 na busca livre na WoS e na rede de autores, para aferir se houve crescimento na quantidade de artigos publicados, como vem-se observando desde 2007.

Outras vertentes de análise poderiam ser reveladoras de características que não foram consideradas para esta pesquisa, como, por exemplo, a área de formação de origem dos autores do *corpus* – a fim de determinar o tipo de abordagem disciplinar e “zonas de influência” dos estudos –, o comportamento das citações do conjunto de artigos, entre outras. A inclusão de mais

autores como ponto de partida na rede de autores, assim como a inclusão de outros tipos documentais na seleção resultariam em um *corpus* maior, e possivelmente mais diverso em termos geográficos ou de área de origem dos autores.

Referências

ARCHAMBAULT, É. *et al.* Benchmarking scientific output in the social sciences and humanities: the limits of existing databases. *Scientometrics*, [Budapest], v. 68, n. 3, p. 329-342, Sept. 2006. Disponível em: <https://link.springer.com/article/10.1007/s11192-006-0115-z>. Acesso em: 2 jan. 2023.

BENOIT, K.; MARSH, M. A relative impact ranking of political studies in Ireland. *The Economic and Social Review*, [s. l.], v. 40, n. 3, p. 269-298, autumn 2009.

BIGDELI, Z. *et al.* Patterns of authors' information scattering: towards a causal explanation of information scattering from a scholarly information-seeking behavior perspective. *Scientometrics*, [Budapest], v. 96, n. 1, July 2013. DOI 10.1007/s11192-012-0885-4. Disponível em: <http://link.springer.com/10.1007/s11192-012-0885-4>. Acesso em: 6 out. 2024.

CHUBIN, D. E.; HACKETT, E. J. *Peerless science: peer review and U.S. science policy*. Albany: State University of New York, 1990.

DAVYT, A.; VELHO, L. A avaliação da ciência e a revisão por pares: passado e presente: como será o futuro? *História, Ciências, Saúde - Manguinhos*, Rio de Janeiro, v. 7, n. 1, p. 93-116, mar./jun. 2000. DOI 10.1590/S0104-59702000000200005. Disponível em: <https://www.scielo.br/j/hcsm/a/QYbkKSZJ4hfBnq4xDsLhDpx/?lang=pt>. Acesso em: 17 fev. 2023.

DONOVAN, C.; BUTLER, L. Testing novel quantitative indicators of research 'quality', esteem and 'user engagement': an economics pilot study. *Research Evaluation*, Oxford, v. 16, n. 4, p. 231-242, Dec. 2007.

ENGELS, T. C. E.; OSSENBLOK, T. L. B.; SPRUYT, E. H. J. Changing publication patterns in the Social Sciences and Humanities, 2000–2009. *Scientometrics*, [Budapest], v. 93, n. 2, p. 373-390, Nov. 2012.

GIMENEZ-TOLEDO, E.; ROMAN-ROMAN, A. Assessment of humanities and social sciences monographs through their publishers: a

review and a study towards a model of evaluation. *Research Evaluation*, Oxford, v. 18, n. 3, p. 201-213, Sept. 2009.

GIMENEZ-TOLEDO, E.; ROMAN-ROMAN, A.; ALCAIN-PARTEARROYO, M. D. From experimentation to coordination in the evaluation of Spanish scientific journals in the humanities and social sciences. *Research Evaluation*, Oxford, v. 16, n. 2, p. 137-148, June 2007.

GIMÉNEZ-TOLEDO, E.; TEJADA-ARTIGAS, C.; MAÑANA-RODRÍGUEZ, J. Evaluation of scientific books' publishers in social sciences and humanities: results of a survey. *Research Evaluation*, Oxford, v. 22, n. 1, p. 64-77, Mar. 2013.

HICKS, D. The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, [Budapest], v. 44, n. 2, p. 193-215, Feb. 1999. Disponível em: <https://link.springer.com/article/10.1007/BF02457380>. Acesso em: 5 ago. 2024.

HICKS, D. The four literatures of social science. In: MOED, H. F.; GLÄNZEL, W.; ULRICH, S. (ed.). *Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of S&T systems*. New York: Kluwer Academic, 2004. p. 473-496.

HICKS, D. *et al.* The Leiden Manifesto for research metrics. *Nature*, [s. l.], v. 520, p. 429-431, Apr. 23 2015. Disponível em: <https://www.nature.com/articles/520429a>. Acesso em: 6 out. 2023.

HUG, S. E.; OCHSNER, M.; DANIEL, H.-D. Criteria for assessing research quality in the humanities: a Delphi study among scholars of English literature, German literature and art history. *Research Evaluation*, Oxford, v. 22, n. 5, p. 369-383, Dec. 2013.

MOED, H. F. *Citation analysis in research evaluation*. Dordrecht: Springer, 2005.

MUGNAINI, R. O Fator de Impacto: sua popularidade, seus impactos e a necessidade de preservação do processo de geração do conhecimento científico. *Revista da Escola de Enfermagem da USP*, São Paulo, v. 50, n. 5, p. 720-721, set./out. 2016. DOI 10.1590/S0080-623420160000600002. Disponível em: <https://www.scielo.br/j/reeusp/a/fwzrqxL6yY9wPZyTTWFFNs/?format=pdf&lang=pt>. Acesso em: 5 out. 2024.

NEDERHOF, A. J. Bibliometric monitoring of research performance in the Social Sciences and the Humanities: a review. *Scientometrics*, [Budapest], v. 66, n. 1, p. 81-100, 2006. Disponível em: <https://link.springer.com/article/10.1007/s11192-006-0007-2>. Acesso em: 5 out. 2024.

OCHSNER, M.; HUG, S. E.; DANIEL, H.-D. Four types of research in the humanities: setting the stage for research quality criteria in the humanities. *Research Evaluation*, Oxford, v. 22, n. 2, p. 79-92, June 2013.

OCHSNER, M.; HUG, S. E.; DANIEL, H.-D. Setting the stage for the assessment of research quality in the humanities. Consolidating the results of four empirical studies. *Zeitschrift Fur Erziehungswissenschaft*, [s. l.], v. 17, p. 111-132, 2014. DOI 10.1007/s11618-014-0576-4. Disponível em: <https://link.springer.com/article/10.1007/s11618-014-0576-4>. Acesso em: 5 out. 2024.

THE ORIGIN of the term 'bibliometrics'. ECOOM, Belgium, 2024. Disponível em: <https://www.ecoom.be/nodes/deoorsprongvandetmbibliometrie/en>. Acesso em: 5 out. 2024.

POPPER, K. R. O progresso do conhecimento científico. In: POPPER, K. R. *Conjecturas e refutações*. Brasília, DF: Ed. UnB, 1980. p. 1-27.

SIVERTSEN, G.; LARSEN, B. Comprehensive bibliographic coverage of the social sciences and humanities in a citation index: an empirical analysis of the potential. *Scientometrics*, [Budapest], v. 91, n. 2, p. 567-575, May 2012. DOI 10.1007/s11192-011-0615-3. Disponível em: <https://link.springer.com/article/10.1007/s11192-011-0615-3>. Acesso em: 5 out. 2024.

SPERA, H. B. *Avaliação da produção científica em Ciências Sociais e Humanas: revisão da literatura recuperada em base de dados e rede de autores*. 2017. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2017. Disponível em: <http://www.teses.usp.br/teses/disponiveis/27/27151/tde-12012018-104739>. Acesso em: 5 out. 2024.

THELWALL, M.; DELGADO, M. M. Arts and humanities research evaluation: no metrics please, just data. *Journal of Documentation*, London, v. 71, n. 4, p. 817-833, 2015. Disponível em: <https://qmro.qmul.ac.uk/xmlui/bitstream/handle/123456789/8423/DELGADO%20-%20Arts%20and%20humanities%20research%20evaluation%3A%20no%20metrics%20please%2C%20just%20data%2C%202015.pdf?sequence=2&isAllowed=y>. Acesso em: 5 out. 2024.

ZUCCALA, A. Quality and influence in literary work: evaluating the 'educated imagination'. *Research Evaluation*, Oxford, v. 21, n. 3, p. 229-241, Sept. 2012.

ZUCCALA, A. *et al.* Altmetrics for the humanities: comparing Goodreads reader ratings with citations to history books. *Aslib Journal of Information Management*, [s. l.], v. 67, n. 3, p. 320-336, 2015.

ZUCKERMAN, H.; MERTON, R. K. Patterns of evaluation in science: institutionalization, structure and functions of the referee system. *Minerva*, London, v. 9, n. 1, p. 66-100, Jan. 1971. Disponível em: <http://www.jstor.org/stable/41827004>. Acesso em: 5 out. 2024.

CAPÍTULO 2

Desafios para o monitoramento do acesso aberto no contexto brasileiro

Ednéia Silva Santos Rocha

Fabiana Andrade Pereira

Introdução

As tecnologias de informação e comunicação – e, em particular, a internet – mudaram fundamentalmente a forma que a ciência e a pesquisa científica são comunicadas. O processo editorial se tornou mais dinâmico, trazendo a possibilidade de uma entrega mundial de conhecimento, aliada à disposição de cientistas e acadêmicos em compartilhar suas pesquisas, e resultou em um movimento global em direção à publicação digital e ao arquivamento segundo as regras da “filosofia aberta”. De acordo com Costa (2006), a filosofia aberta define o movimento que preza um novo processo de comunicação científica, que se utiliza de ferramentas e métodos que compreendem os *softwares* abertos, os arquivos abertos e o Acesso Aberto (AA).

O AA, em inglês Open Access (OA), é um conjunto de iniciativas mundiais para a disponibilização *on-line* e sem limitações das informações científicas publicadas (Organização das Nações Unidas para a Educação, a Ciência e a Cultura, 2015). O movimento iniciou com a proposta de compartilhamento de artigos de periódicos revisados por pares. Entretanto, devido ao impulso da Ciência Aberta – Open Science (OS) –, atualmente, conta-se com diversos tipos de documentos, informações e dados que podem ser compartilhados, com o objetivo de promover uma ampla divulgação de resultados de pesquisas, aumentando o seu uso, a sua aplicação e o seu impacto.

O AA tem como premissa que o conhecimento científico, principalmente aqueles que receberam financiamento público, pertence a toda sociedade e, portanto, deve estar disponível para todos os interessados, sem

barreiras de acesso. Apesar do amplo sentido que o AA envolve, à medida que o movimento cresce, também aumenta a confusão sobre o que a iniciativa realmente significa, particularmente no que diz respeito ao uso irrestrito do conteúdo que o verdadeiro AA permite (Maccallum, 2007).

Com isso, desde a Declaração de Bethesda, e de forma ampla, o AA trata de um posicionamento político frente à utilização e à reutilização de publicações e demais informações científicas, mediante alguns critérios estipulados por suas licenças – como a Creative Commons Attribution, por exemplo – e desde que o autor seja corretamente citado. Desafortunadamente, nem sempre o acesso irrestrito significará gratuidade de acesso, porque de qualquer forma haverá custos nos processos editoriais, seja para que os autores publiquem e ofereçam acesso gratuito – por Via Dourada ou Acesso Ouro –, ou no tempo despendido para disponibilizá-los em repositórios abertos, após o tempo de embargo exigido – por Via Verde ou Acesso Verde (Marques, 2016). O verdadeiro valor que o AA traz para a comunidade científica se deve à autonomia em relação aos processos editoriais da sua própria comunicação.

De qualquer forma, o panorama da publicação científica mudou consideravelmente e abriu o mercado para formas alternativas de distribuição de literatura científica. Além dos benefícios referentes à eficiência no processo de comunicação científica, em que a entrega de conteúdo digital se tornou relativamente curta (Laakso *et al.*, 2011).

Weitzel (2014, p. 66) aponta que o AA passou a exercer grande impacto na “estrutura de poder econômico que envolve o sistema de publicação científica”. Isso posto, por meio do AA, a comunidade científica se manifesta intrinsecamente contra os modelos tradicionais de disseminação, dominados pela indústria editorial de periódicos científicos indexados e arbitrados, que já não correspondem aos ideais da comunicação científica: primeiramente, devido ao alto custo despendido para publicação e à assinatura dos periódicos científicos por instituições de pesquisas; posteriormente, deve-se à demora, à dificuldade de acesso e de visibilidade dos resultados das pesquisas para todos que não podem assiná-las.

O “novo paradigma” da editoração proposto pelo AA (Kuramoto, 2008, p. 865) preconiza o uso de novos recursos de baixo custo e de grande capacidade de promoção da informação científica, muitas vezes resultantes

das pesquisas financiadas com recursos públicos, para serem armazenadas e disponibilizadas em repositórios institucionais. Dessa forma, devido à perspectiva de rapidez na divulgação e na redução de custos, o movimento encontrou apoio e engajamento de instituições e governos, uma vez que trouxe a possibilidade de oferecer acesso e visibilidade de disciplinas ou áreas temáticas emergentes, como para os países periféricos, social e historicamente negligenciados por exclusão cognitiva¹, que não fazem parte do fluxo internacional e central da informação científica. Assim, para os cientistas e pesquisadores de toda a América Latina, as iniciativas de AA significaram a obtenção de oportunidades no estabelecimento de infraestruturas necessárias para ampliar a divulgação da produção científica (Costa; Leite, 2016).

Embora a situação do AA tenha melhorado consideravelmente com o passar desses mais de 20 anos de existência do movimento, nota-se que a questão da indexação de periódicos ainda representa um grande desafio. Torna-se, também, um grande desafio escrever, minuciosamente, a totalidade das abordagens e ações nacionais e internacionais em prol do AA. Nesse sentido, como destaca Kuramoto (2008, p. 866), se as agências de fomento, governos, instituições de pesquisa e os próprios pesquisadores foram os principais atores na mudança do “paradigma da comunicação científica tradicional”, cabe a este capítulo investigar como está o apoio e o engajamento desses organismos no atual estado do movimento de AA. Dessa maneira, este capítulo descreve, sucintamente, de que maneira as políticas ou as diretrizes, sobretudo das agências de fomento, influenciam as universidades e, conseqüentemente, este *modus operandi* das comunidades científicas em ações voltadas à adesão e à ampliação do AA. Além disso, busca-se observar os desafios referentes ao monitoramento do AA por esses órgãos, para verificar se o AA está se expandindo e quais modelos de AA estão sendo implementados.

1 De acordo com Kuramoto (2008, p. 866), “o compartilhamento desse conhecimento, que foi essencialmente produzido no hemisfério Norte, é uma condição necessária, mas não suficiente, para diminuir as desigualdades existentes no mundo. A combinação desses dois fatores, concentração do conhecimento no hemisfério Norte e os custos extorsivos das publicações científicas, proporciona o fenômeno que está sendo conhecido como exclusão cognitiva. Incluem-se, nesse fenômeno, aqueles que não possuem acesso à informação científica, especialmente os pesquisadores de países em desenvolvimento e não desenvolvidos”.

Todavia, para entender esses aspectos, é necessário observar, revisitar o histórico e analisar as primeiras iniciativas e as características específicas do AA.

Histórico e princípios que alicerçaram o movimento AA

As primeiras iniciativas de compartilhamento gratuito de artigos científicos são atribuídas à CiteSeer² que, na década de 1980, foi a primeira biblioteca digital com mecanismo de pesquisa que oferecia indexação e *links* de citações automatizados por indexação autônoma de citações.

No início da década de 1990, evidencia-se também o ArXiv³, o primeiro repositório de pré-prints eletrônicos de artigos científicos idealizado pelo físico Paul Ginsparg, no Los Alamos National Laboratory (Van de Sompel; Lagoze, 2000). O repositório foi impulsionado pelas demandas da comunidade científica da área de física, que necessitava de uma plataforma de compartilhamento das pesquisas entre os pares. Na área de física, o compartilhamento de informações entre os cientistas era uma prática já utilizada por meio de cartas ou de correio eletrônico. Esses arquivos eram chamados de *open archives* – uma das primeiras denominações do que conhecemos atualmente como repositórios pré-prints e pós-prints. Por volta de 1994, Stevan Harnad apresentou uma proposta para as publicações eletrônicas, publicada posteriormente pela Association of Research Libraries⁴.

Segundo Laakso e demais autores (2011), o movimento ocorreu em três fases, que cronologicamente destacam-se em:

1. 1993-1999: anos do “pioneirismo”. Os periódicos AA foram quase exclusivamente fundados por acadêmicos ou grupos de acadêmicos e publicados em plataformas tecnicamente simples. O “modelo de negócios” comumente consistia em trabalho voluntário combinado com a possibilidade de usar os servidores *web* da universidade gratuitamente.

2 Mais informações sobre o CiteSeerX disponíveis em: <https://csxstatic.ist.psu.edu/home>.

3 Mais informações sobre o ArXiv disponíveis em: <https://arxiv.org/about>.

4 Scholarly journals at the crossroads: a subversive proposal for electronic publishing. Disponível em: <https://www.arl.org/resources/scholarly-journals-at-the-crossroads-a-subversive-proposal-for-electronic-publishing/>.

O modelo tem semelhanças com o modelo tradicional de produção de *software* de código aberto;

2. 2000-2004: anos de “estabilidade no crescimento”. Surgiram novos modelos de negócios para executar operações de AA em uma escala mais ampla. Surge a BioMedCentral, que é um serviço destinado a autores para publicação de artigos de AA em periódicos de qualidade e alto fator de impacto. Curiosamente, foi financiada por capital de risco e, mais tarde, comprada pela Springer em 2008. Public Library of Science foi capaz de estabelecer um sistema de gestão de periódicos de alta qualidade com um investimento financeiro inicial substancial. Neste período, a digitalização geral de periódicos impressos estabelecidos foi rápida e vários outros periódicos começaram a usar os serviços da HighWire Press⁵;

Em outros países e regiões, os portais de publicação AA – ou seja, a Scientific Electronic Library Online (SciELO) na América Latina e o J-STAGE no Japão – permitiram que os periódicos que atendiam aos requisitos tivessem uma exposição global muito mais ampla para seus artigos.

Em 2004, as editoras convencionais começaram a experimentar o modelo híbrido – por exemplo, a Springer Open Choice –, que permitia aos autores de artigos em periódicos tradicionais por assinatura abrirem seus artigos mediante o pagamento de uma taxa. Foi também nesse período que a *advocacy* do AA se tornou muito mais visível. Várias declarações na *web* – por exemplo, a Budapest Open Access Initiative (BOAI) – estabeleceram os princípios centrais do AA, e surgiram conferências e até séries de conferências dedicadas ao AA.

3. 2005-2009: anos de consolidação. Surgem, neste período, cada vez mais a infraestrutura de suporte à publicação AA. O Directory of Open Access Journals (DOAJ) se tornou o principal índice de periódicos AA, oferecendo também possibilidades de arquivamento de longo prazo por meio de um acordo com a Dutch National Library. Um número cada vez maior de periódicos adotou e está adotando o *software* gratuito

5 Mais informações sobre o HighWire disponíveis em: <https://www.highwirepress.com/>.

Open Journal Systems. Os acordos de licenciamento adequados para periódicos AA, principalmente versões de licenças Creative Commons, também ganharam aceitação crescente. Nos últimos anos, várias novas editoras comerciais AA entraram no mercado. Algumas das grandes e antigas editoras científicas começaram a oferecer periódicos de AA financiados por pequenas taxas para o autor. A Open Access Scholarly Publishers Association foi fundada para ajudar os editores AA e para definir padrões de qualidade. Autores que pensam em publicar em periódicos AA também são ajudados pelo fato de que muitos órgãos de fomento científico permitem que os custos de publicação AA sejam incluídos nos orçamentos de pesquisa, e que algumas universidades reservaram fundos para esse fim.

Nesse cenário, no Quadro 1, a seguir, são destacadas concisamente as declarações da comunidade científica em prol do AA.

Quadro 1 – Declarações que apoiaram o AA

Iniciativa	Ano	Descrição
BOAI	2002	Considerou o AA como o acesso totalmente irrestrito e gratuito à produção científica revisada por pares e disponível no meio eletrônico, rompendo qualquer impedimento de acesso a esta literatura científica.
Bethesda	2003	Teve o intuito de assegurar o acesso livre à literatura científica, com a proposta de que os autores possam conceder os direitos e estipular as condições sobre a disponibilização das suas produções.
Berlin	2003	Apresentou duas diretrizes para o AA, sendo elas: 1) a concessão completa dos direitos dos autores, com imposição de uso responsável; e 2) a descrição do formato e a normalização das obras para disponibilização em repositórios, para garantir o acesso irrestrito.

Fonte: adaptado de Andrade e Muriel-Torrado (2017, p. 2-3).

Como pode-se observar, a proposta da BOAI (2002) ressalta que artigos de AA são aqueles cuja disponibilidade é pública e gratuita na internet, permitindo que qualquer usuário leia, baixe, copie, distribua, imprima, pesquise ou crie *links* para os textos completos desses artigos, utilizando essas publicações para quaisquer outros fins lícitos, sem barreiras financeiras, jurídicas ou técnicas. Prevalecendo os direitos autorais, pois os autores são

os responsáveis pela integridade dos conteúdos e têm o direito de serem devidamente reconhecidos e citados.

Essa definição destaca dois canais distintos de acesso: (I) acesso humano para ler, baixar e reutilizar o texto completo de artigos publicados e (II) acesso à máquina para rastrear, indexar, ou analisar o conteúdo dos artigos.

A BOAI também propõe dois caminhos operacionais para acesso por meio de periódicos de AA e autoarquivamento em repositórios, posteriormente denominados Gold OA e Green OA (Bailey Junior, 2005).

A Declaração de Budapeste ressalta duas importantes estratégias de viabilização do AA, que seriam: a Via Verde e Via Dourada. Ao longo do tempo surgiram outras formas de concretização do AA determinadas pela versão que o artigo é publicado, quando ou onde é publicado, criando assim tipologias de acesso complexas:

Quadro 2 – Estratégias para o AA

Acesso Verde (Via Verde ou <i>Green Road OA</i>)	Versões autoarquivadas de um manuscrito. Aqui, a responsabilidade recai sobre os autores da publicação, ou colegas institucionais, como membros da equipe da biblioteca central, que supervisionam o depósito do documento em um repositório. Esta versão do documento pode não corresponder à versão final do editor.
Acesso Ouro (Via Dourada ou <i>Golden Road OA</i>)	Isto se refere a periódicos que publicam todos os seus manuscritos em AA, independentemente do modelo de negócios que seguem – por exemplo, patrocinado publicamente, ou a publicação é feita por periódicos de editores comerciais que cobram Article Processing Charge (APC).
Acesso Híbrido (Via Híbrida ou <i>Hybrid Road OA</i>)	Periódicos com acesso pago (não AA) tornam publicações específicas abertamente acessíveis, geralmente depois que o autor paga uma taxa, alegando uma suposta necessidade de contabilizar perdas potenciais derivadas de taxas de assinatura.
Acesso Bronze (Via Bronze ou <i>Bronze Road OA</i>)	Este tipo de AA foi sugerido pela primeira vez por Piwowar e demais autores (2018) e refere-se a artigos de leitura gratuita, disponibilizados por editores, sem uma menção explícita a qualquer licença AA. Este AA não está sujeito às condições de direitos autorais definidas como AA – ou seja, eles não garantem acesso gratuito perpétuo.
Acesso Diamond (Via Diamante ou <i>Diamond Road OA</i>)	Este modelo de AA é caracterizado por não cobrar taxas dos autores (APCs) para processamento e publicação de artigos, e não há custos para os leitores. O financiamento vem de organizações acadêmicas, sociedades científicas, doações e outras fontes que cobrem os custos operacionais.

Fonte: elaborado pelas autoras.

A Via Verde é considerada o modelo ideal para promoção do AA, e se materializa por meio da disponibilização dos artigos, avaliados por pares, em repositórios abertos (Björk *et al.*, 2013).

Apesar do crescimento significativo do número de trabalhos de pesquisa disponíveis por meio do AA, principalmente por meio do autoarquivamento de autores em repositórios institucionais, estimava-se que apenas 20% do número de artigos publicados anualmente são de AA. No entanto, em 2020, de acordo com a plataforma Dimensions⁶, o AA superou a publicação de artigos em periódicos de acesso restrito. De acordo com Hook (2021), esse resultado pode estar relacionado a duas importantes transformações: a primeira se refere à dicotomia entre pesquisa e dados, pois o Big Data modificou a forma de se obter dados, analisar os resultados e comunicar as pesquisas, e, a segunda se deve à transformação que ocorreu no ambiente das instituições para se tornarem mais transparentes e responsáveis. No entanto, mudanças estruturais e culturais levam tempo para se estabelecer, e requerem mobilização contínua da comunidade científica para manter as propostas em movimento.

Toda essa diversidade de conceitos relacionados ao AA acaba dificultando o monitoramento eficiente desse movimento, principalmente porque não existe uma forma padronizada de identificar esses manuscritos por meio dos metadados bibliográficos.

Alguns membros de comunidades científicas com financiamento limitado não possuem acesso a uma grande parcela da literatura científica, o que pode afetar a qualidade do desenvolvimento das pesquisas. De acordo com Dunn, Coiera e Mandl (2014), apenas 13% de todos os artigos indexados no PubMed estavam disponíveis gratuitamente no PubMed Central (PMC). Para minimizar o problema, abordagens foram empregadas para melhorar o acesso à literatura. Uma delas foi o modelo de AA híbrido, em que os autores tiveram a opção de tornar os artigos em AA no momento da aceitação do manuscrito. Isso estava sujeito ao pagamento de taxas de processamento do artigo (APCs). Infelizmente, APCs geralmente custam caro e

6 Dimensions publications. Disponível em: <https://app.dimensions.ai/discover/publication>.

são economicamente inviáveis para a maioria dos pesquisadores, principalmente em áreas do conhecimento que recebem menos fomento científico.

Papel das universidades no apoio ao AA

A estratégia mais comum utilizada pelas universidades para promover o AA tem sido manter repositórios institucionais e introduzir mandatos/portarias que incentivem seus pesquisadores a depositarem pré-prints ou pós-prints de suas publicações (Harnad, 2007; Harnad *et al.*, 2008; Robinson-Garcia; Costas; Van Leeuwen, 2020). Há também evidências de instituições que promovem publicações em AA, patrocinando os custos derivados das taxas de processamento de artigos de periódicos abertos (Gorraiz; Wieland, 2009; Gorraiz; Wieland; Gumpenberger, 2012).

Na maioria dos casos, as instituições enfrentam o desafio de determinar o sucesso de tais iniciativas e monitorar o cumprimento de seus pesquisadores das diretrizes nacionais e internacionais de AA (Robinson-Garcia; Costas; Van Leeuwen, 2020). Iniciativas como o *ranking* de repositório de AA oferecem informações parciais sobre compartilhamento/disponibilidade de AA no nível institucional, visto que apenas fornecem detalhes sobre a relação de documentos armazenados em repositórios institucionais independentemente do resultado geral da pesquisa (Aguillo *et al.*, 2010).

Até recentemente, não havia mais do que estimativas quanto à quantidade de publicações em AA. No entanto, o desenvolvimento de plataformas como o CrossRef, DOAJ, ou mesmo o Google Scholar, com avanços computacionais em *web scrapping*, levaram a uma infinidade de análises em grande escala para identificar empiricamente a literatura em AA (Archambault *et al.*, 2014; Martín-Martín *et al.*, 2018; Piwowar *et al.*, 2018; Van Leeuwen; Tatum; Wouters, 2018).

Robinson-Garcia, Costas e Van Leeuwen (2020) ressaltam que esses estudos relatam que cerca da metade da literatura científica está disponível gratuitamente, mas apontam para o aumento da disponibilidade de publicações que não aderem estritamente ao que é considerado AA.

Explorar as infraestruturas que facilitem o alcance aos conteúdos abertos é fundamental para promoção do livre acesso às publicações científicas, pois

existem fontes ilegítimas de acesso que podem não ser consideradas adequadas por não respeitarem os direitos intelectuais dos autores ou editoras, como o Sci-Hub. Gadhvi, Srivastav e Sharma (2020) destacam algumas fontes de informação legitimadas para acesso a trabalhos científicos, dentre essas destaca-se a extensão Unpaywall, que lhe permite encontrar versões disponíveis gratuitamente. A aplicação foi desenvolvida pelo Impactstory⁷, e usa a Interface de Programação de Aplicativos (API) para encontrar a variante em livre acesso dos artigos legalmente depositados (Aventurier, 2017). A ferramenta utiliza Open Access Digital Object Identifier (oaDOI) para localizar os artigos.

A Open Access Button (OAB)⁸ é uma plataforma lançada pelos criadores Joseph McArthur e David Carroll e está disponível como um *plug-in* de *site* e navegador. Ela pesquisa agregadores de AA, como CORE, BASE e oaDOI para cópias legítimas de documentos. O Kopernio⁹ é um *plug-in* de navegador *web* que pesquisa em diversas bases, entre elas: PubMed, Web of Science (WoS), Google Scholar, JSTOR e outras fontes de AA válidas para cópia do papel. A ferramenta pertence à Clarivate AnalyticsTM e tem integração com a WoS, que, de acordo com Else (2018c), favorece a divulgação da base de dados. A ferramenta também pode se integrar com credenciais de *login* institucionais e facilitar o acesso ao papel.

Outra ferramenta que se destacam, é a Lazy Scholar¹⁰, uma extensão de navegador *web* que utiliza fontes como Google Scholar, PubMed, Europe PMC, DOAI.io e Dissemin.in para cópias legítimas gratuitas do papel, além da ArXiv¹¹, que é uma rede gratuita de repositórios e distribuição de artigos de periódicos mantidos pela Universidade de Cornell (Gadhvi; Srivastav; Sharma, 2020).

7 Heather Piwowar e Jason Priem são fundadores da Impactstory, que é um *site* de código aberto que ajuda os pesquisadores a explorar e compartilhar o impacto *on-line* de suas pesquisas. Sendo uma empresa sem fins lucrativos, criaram também o Unpaywall, oaDOI e outras ferramentas para melhorar a comunicação científica. A organização é financiada pela National Science Foundation e Alfred P. Sloan Foundation (Piwowar *et al.*, 2018).

8 Open Access Button. Disponível em: <https://openaccessbutton.org/>.

9 Sobre o Kopernio. Disponível em: <https://kopernio.com/>.

10 Lazy Scholar. Browser extension. Disponível em: <http://www.lazyscholar.org/>.

11 ArXiv.org. Disponível em: <https://arxiv.org/>.

Nesse cenário, os identificadores persistentes são essenciais para o pleno funcionamento dessas aplicações de rastreabilidade de publicações em AA, pois o Digital Object Identifier (DOI), representa um padrão cujo objetivo é identificar documentos virtuais, relevante para auxiliar na rastreabilidade para localização de qualquer item documentário na *web*, facilitando o acesso e a autenticação de documentos.

Papel das agências de fomento no apoio ao AA

O impulso para o AA tem crescido à medida que várias agências de financiamento e instituições em todo o mundo implementaram políticas científicas direcionadas ao AA. Múltiplas agências nacionais de financiamento juntaram as mãos com a proposta de que qualquer pesquisa financiada por elas deveria ter AA imediatamente. No entanto, um fator crítico para o sucesso das políticas AA é a disponibilidade da infraestrutura necessária para implementar tais políticas.

A União Europeia (UE), por exemplo, tem se esforçado para promover o acesso livre e para aumentar os benefícios do público por meio do investimento na pesquisa financiada no âmbito do programa de pesquisa Horizon 2020¹². Para este fim, foi desenvolvida a OpenAIRE¹³ para gerenciar e monitorar os resultados do financiamento da UE em pesquisas que apoiam as políticas de AA.

Robinson-Garcia, Costas e Van Leeuwen (2020) destacam que a implementação de políticas que promovam a adoção de uma cultura de OS) deve ser acompanhada de indicadores que permitam monitorar a adoção de tais políticas e seus potenciais efeitos nas práticas de publicação e compartilhamento de pesquisas. A noção de ciência aberta remonta ao século XX (David, 2008), mas tem recentemente tido relevância à medida que a UE introduziu políticas científicas em seus programas de pesquisa (Moedas, 2015). Em diferentes diretrizes estabelecidas para alcançar o AA, tornou-se

12 What is Horizon 2020? Disponível em: <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>.

13 OpenAIRE – Open Access Infrastructure for research in Europe. Project website. Disponível em: <https://www.openaire.eu/>.

um dos primeiros marcos as iniciativas como o Plano S (Else, 2018a, 2018b) ou o Open Science Monitor¹⁴ que se trata de uma ferramenta de monitoramento da adoção do AA utilizada pela European Commissions, exemplificando esses esforços e a priorização do AA para essas agências.

Outra ação que impulsionou o movimento do AA está relacionada ao cOAlition S¹⁵ ou Plano S, que aborda um acordo entre diversas agências de fomento público que priorizam a publicação das pesquisas científicas em AA para leitura e *download*. Essa iniciativa é um importante incentivo para o movimento, pois busca acelerar a transição para o AA total e imediato às publicações científicas. Apesar dos esforços da UE em prol do AA, dados do The Society of College, National and University Libraries (Sconul) –, ressaltam que 37% dos trabalhos dos cientistas do Reino Unido são publicados em AA.

Embora tenha sido uma iniciativa bem-vinda, foi restrita a apenas alguns países, isso porque, – somente em alguns houve a elaboração de políticas científicas para disponibilizar em AA todas as pesquisas financiadas com recursos públicos. Vale ressaltar que várias propostas encontram-se pendentes de implementação, pois se observa que o monitoramento e o apoio efetivo dos organismos de fomento ainda são ineficazes. Apesar das medidas citadas, a proporção de literatura científica livremente acessível ao público em geral permanece bastante pequena.

A Science Europe ressalta que o sistema de recompensas nos campos científicos necessita estimular a cultura científica para publicação em AA. Iniciativas como a Declaração de São Francisco sobre Avaliação de Pesquisa (Dora), o Manifesto de Leiden, e a Matriz de Avaliação de Carreira do Open Science (OS-CAM) são passos em direção a isso, além do Fórum de Métricas Responsáveis de Pesquisa.

A importância das principais agências de fomento nacionais no apoio ao AA

14 European Commission. Open Science Monitor. Disponível em: https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor_en.

15 cOAlition S - Plan S. Disponível em: <https://www.coalition-s.org/>.

No Brasil, a maior parte do investimento em Ciência, Tecnologia e Inovação (CT&I) advém de recursos públicos. A descentralização das atividades administrativas no Brasil¹⁶ é uma das características que mais afeta as suas políticas públicas. Trata-se de um processo complexo, pois funciona por distribuições de funções executivas e decisórias em diferentes níveis. Como consequência, as ações de financiamento e apoio à pesquisa no país ocorrem em diversas instâncias, envolvendo atores institucionais das esferas públicas e privadas, além de órgãos municipais, estaduais e federais. Conforme destacado pelo Centro de Gestão e Estudos Estratégicos (2010), a relevância dessa característica se intensifica diante das consideráveis diferenças regionais e da singularidade do Brasil como a única federação trina global, onde os municípios têm autonomia federativa, possuindo direitos e obrigações completos frente à federação. O mesmo documento também menciona que a descentralização das políticas públicas de fomento em CT&I são recentes, avançando principalmente a partir dos anos 2000, assim responsável pela atual configuração do Sistema Nacional de Ciência, Tecnologia e Inovação (SNCTI) (Centro de Gestão e Estudos Estratégicos, 2010).

Dessa forma, entende-se que o apoio e o financiamento da pesquisa em CT&I no Brasil ocorre por diversas instâncias, em um sistema composto por diferentes instituições de fomento à pesquisa e outras entidades que são vinculadas, direta ou indiretamente, ao governo federal e, em muitas vezes, operando em parceria. A descentralização administrativa gera a descentralização das políticas e programas de fomento em CT&I no Brasil, destacando os impactos e as desigualdades entre os estados, cujos dispêndios

16 Descentralização se refere à distribuição de funções entre diferentes níveis de governo – federal, estaduais e municipais – e entre esses níveis e instituições do setor privado. O processo de descentralização possui duas dimensões analíticas: a descentralização da responsabilidade pela execução das políticas ou dos programas e a descentralização da autoridade decisória sobre essas políticas. No contexto do processo de descentralização, as instituições do setor privado não se referem às empresas tratadas isoladamente e que, em alguns programas CT&I, são os atores finalísticos das ações dos órgãos de fomento – federais, estaduais ou municipais –, mas sim às instituições que, na gestão desses programas, podem compartilhar competências e autoridade decisória com esses órgãos, a exemplo de associações ou sociedades empresariais; federações de indústrias (Centro de Gestão e Estudos Estratégicos, 2010, p. 12).

serão distintos e dependerão do alinhamento das políticas entre as instâncias governamentais (Centro de Gestão e Estudos Estratégicos, 2010).

A estruturação institucional do apoio e fomento em CT&I decorre da articulação entre as seguintes instâncias (Centro de Gestão e Estudos Estratégicos, 2010, p. 37):

Ao nível federal: o Fundo Nacional de Desenvolvimento Científico e Tecnológico (FNDCT)¹⁷, criado em 1969, é o principal instrumento público de financiamento de ciência, tecnologia e inovação do Brasil. A Financiadora de Estudos e Projetos (Finep) exerce a função de Secretaria-Executiva – responsável pelas atividades de natureza administrativa, orçamentária, financeira e contábil – do FNDCT desde 1971 e ratificado na Lei nº 11.540 de 2007 (BRASIL, 2021). Desde então, a Finep, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (Capes) passaram a compor a principal tríade federal de agências públicas de fomento à CT&I no Brasil;

Ao nível estadual: os fundos estaduais de CT&I, Fundações de Amparo à Pesquisa (FAPs), algumas vezes com o intermédio do Conselho Nacional das Fundações Estaduais de Amparo à Pesquisa (Confap) – organização cujo objetivo é promover uma melhor articulação dos interesses das 26 FAPs. Além disso, articulam-se em conjunto com o governo federal, por meio de programas de fomento, compartilhados entre os estados e instituições do setor privado.

As principais agências de fomento brasileiras e demais entidades vinculadas ao apoio e ao financiamento da pesquisa em CT&I no Brasil vêm reconhecendo a necessidade de monitorar o retorno sobre os projetos de pesquisas que obtiveram apoio público. Mugnaini, Igami e Kryzanowski (2022) destacam que as agências de fomento do Brasil exercem uma coerção crucial para que os resultados das pesquisas apoiadas com recursos públicos estejam também disponíveis para todos. Atualmente muitos financiadores

17 Sobre o FNDCT. Disponível em: <https://antigo.mctic.gov.br/mctic/opencms/fundos/fndct/paginas/sobre.html>.

exigem que as publicações resultantes de seu apoio sejam disponibilizadas ao público imediatamente ou após o período de embargo.

De acordo com Freire (2011), tais exigências exercem impactos positivos, pois influenciam os pesquisadores nos processos de autoarquivamento de artigos em repositórios institucionais, como também influenciam as editoras científicas a encontrar meios de adequar seus negócios frente à nova realidade do AA. Mugnaini, Igami e Kryzanowski (2022) observam que as bases de dados, por exemplo, já estão se adequando e disponibilizando metadados para “menção à fonte de financiamento da pesquisa”. Por meio dessas marcações que as agências de fomento podem verificar as atividades de pesquisa por elas financiadas, embora essa prática ainda esteja em fase de consolidação, cuja obrigatoriedade do preenchimento dessas informações vem sendo reforçada nos últimos anos, destacam os autores.

O CNPq, atualmente vinculado ao Ministério da Ciência, Tecnologia e Inovações (MCTI), é responsável por diversas ações em andamento em prol do crescimento da OS no Brasil, sobretudo do acompanhamento e da participação de marcos regulatórios, destacam Rezende e Abadal (2020). Esses marcos regulatórios são importantes por promoverem um posicionamento da entidade acerca da OS e do AA nas temáticas de sensibilização, dos procedimentos operacionais e de regulamentação, como forma de proporcionar novas possibilidades para o avanço científico e tecnológico (Parker; Santos, 2019; Tonini; Santana; Scherrer, 2020).

Desde 2005, o CNPq se manifesta ativamente a favor do AA, oferecendo apoio às bibliotecas eletrônicas e repositórios de publicações científicas, fazendo acordos institucionais para interoperabilidade entre sistemas, permitindo acesso às informações de pesquisadores, grupos de pesquisas e publicações da Plataforma Lattes com bases de dados como SciELO e Scopus, além da inclusão de critérios que privilegiam o AA em chamadas do Programa Editorial¹⁸. O programa encoraja os beneficiários de projetos e bolsas na publicação de trabalhos seguindo os princípios do AA, desta-

18 O objetivo do Programa Editorial do CNPq é apoiar e incentivar a editoração e a publicação de periódicos científicos brasileiros em todas as áreas de conhecimento, priorizando o apoio às revistas divulgadas por meio eletrônico, na internet, em modo de acesso aberto, ou de forma impressa/eletrônica simultaneamente (Tonini; Santana; Scherrer, 2020, p. 110).

cando que as produções científicas oriundas de seu financiamento são propriedade de seus autores e não necessariamente dos periódicos e órgãos de publicação¹⁹. O CNPq se manifesta a favor do AA fazendo recomendações, pois “entende que a publicação científica, seja por meio de periódicos, livros ou anais de eventos, [...] constitui a principal ferramenta da divulgação e disseminação da produção intelectual” destacam Tonini, Santana e Scherrer (2020, p. 204).

Outro aspecto a ser observado foi a incorporação obrigatória às chamadas do CNPq da possibilidade de utilização de recursos para pagamento de custos de publicação no modelo de AA (Tonini; Santana; Scherrer, 2020, p. 210-211).

Verificando a influência das recomendações e decisões do CNPq sobre o *modus operandi* dos pesquisadores bolsistas de produtividade 1A, o estudo desenvolvido por Leta, Costa e Mena-Chalco (2017, p. 4) verificou que a produção em periódicos em AA obteve um aumento significativo no período de 2000-2015, apontando que esses pesquisadores têm optado por publicações neste formato para divulgação das pesquisas.

Assim, a decisão dos pesquisadores bolsistas de produtividade do CNPq em favorecer o formato de AA para suas publicações reflete não apenas a influência das políticas do CNPq, mas também um alinhamento com as tendências globais de comunicação científica, visando maximizar o alcance e o impacto de suas pesquisas. Embora sejam perceptíveis os esforços do CNPq no que tange a essas questões, ainda há muito no que avançar ao estímulo à OS e ao AA.

No que concerne à Capes, atualmente vinculado ao Ministério da Educação, lançou em novembro de 2000 o Portal de Periódicos Capes, no qual passou a “centralizar e otimizar a aquisição de conteúdo científico, por meio da negociação com editores e distribuidores”, observam Melo e Café (2018, p. 311). Os autores também destacam que, desde sua criação, o Portal de Periódicos oferece acesso às informações científicas pagas, visando aliviar as tensões de aquisição e as dificuldades de acesso pelas instituições de ensino superior e pesquisa e pelas agências de fomento à pesquisa. Por conta

19 Informação presente na Ata da 133ª Reunião do Conselho Deliberativo do CNPq.

do surgimento do movimento de AA, o portal passou a ofertar visibilidade às produções científicas em AA, lançando em 2006 o Portal de Acesso Livre da Capes²⁰.

Em 2007, a Capes com o CNPq abriu uma seleção pública conjunta para apoiar propostas de editoração e publicação de periódicos científicos brasileiros, em que teriam prioridade as propostas de periódicos divulgados na internet com AA (Brasil, 2007). Em 2017, assim como a SciELO, a Capes também assinou a manifestação de interesse²¹ ao movimento Open Access 2020 Initiative (OA2020)²², em que se compromete a favor da aceleração do AA em território brasileiro, por meio do apoio ao desenvolvimento de plataformas de publicação em AA, como arquivos e repositórios; assim como o incentivo aos pesquisadores para a publicação em periódicos envolvidos com o movimento.

No nível estadual são destacadas as FAPs estaduais, órgãos de fomento à pesquisa que estão ligados às Secretarias de Desenvolvimento Econômico, Ciência, Tecnologia e Inovação do governo de cada estado²³. Elas têm como objetivo viabilizar os recursos estaduais e federais para o desenvolvimento da pesquisa nas diferentes áreas das CT&I. As FAPs integram o Sistema Nacional de Ciência e Tecnologia (SNCT), conjuntamente ao CNPq e à Finep, e exercem um papel significativo nas definições das políticas científico-tecnológicas nacional (Menezes, 2001).

A primeira FAP do Brasil foi a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), fundada em 1962. A Fapesp serviu de modelo para a estruturação das demais fundações dos estados brasileiros. “Cada FAP tem sua constituição e regimento próprios, cujo orçamento é ligado à receita estadual”, menciona Alves (2018, p. 8) e, para aumentar as oportunidades

20 O Portal de Acesso Livre disponibilizava para professores e alunos da pós-graduação de todo o país parte do conteúdo do Portal da Capes – 1.050 revistas científicas brasileiras e internacionais, resumos de bases referenciais e de patentes, resumos de 175 mil teses de dissertação, entre outras bases de dados (Brasil, 2022).

21 *Expression of interest in the large-scale implementation of open access to scholarly journals*. Disponível em: <https://oa2020.org/mission/>.

22 Open Access 2020. *Learn More*. Disponível em: https://oa2020.org/learn_more/.

23 São 26 fundações distribuídas em quase todos os estados da Federação – a exceção é Roraima – incluindo o Distrito Federal (Alves, 2018).

para pesquisadores e instituições de pesquisa, constantemente conduzem acordos com agências nacionais como o CNPq, a Capes, a Finep e diversos organismos internacionais. Após o CNPq, a Fapesp é o órgão de fomento que mais oferece apoio às pesquisas em CT&I no Brasil. A Fapesp, por meio da sua atuação seguindo tendências mundiais, busca seguir diretrizes e orientações internacionais para que a participação da comunidade científica paulista alcance maior visibilidade nacional e internacional (Shintaku; Lanne, 2020).

O projeto SciELO, lançado em 1998, quatro anos antes da BOAI, com parceria do Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (Bireme) (Fausto, 2013), marca o pioneirismo da Fapesp no apoio ao AA (Cruz, 2019). A SciELO também se tornou um exemplo de qualidade e precisão editorial de publicações em AA e alavancou as publicações nacionais e de demais países da América Latina.

Ortellado (2008) aponta que, dada a falta de um mercado significativo de revistas científicas no Brasil, os financiamentos para os periódicos nacionais vieram predominantemente de fontes governamentais. Nesse contexto, a transição para o modelo de AA tornou-se uma decisão estratégica tomada pelas entidades de fomento à pesquisa, iniciando com a Fapesp e seguindo com o CNPq.

Conforme se observou em alguns meios de divulgação científica (Fundação de Amparo à Pesquisa do Estado de São Paulo, 2021; Hey, 2017; Marques, 2019; SciELO, 2013), formalmente, desde 2008, a Fapesp vem buscando adotar algumas políticas que favorecem o AA. Essa primeira política sugere, de forma mais geral, que é de interesse da fundação que os artigos publicados originados de projetos financiados, devam ser depositados em repositórios institucionais. Com o passar dos anos, surgiram diversos manifestos nacionais e internacionais de apoio ao AA e a importância da SciELO cresceu significativamente, fazendo com que outras iniciativas relativas ao AA fossem observadas pela Fapesp. A criação e estruturação de repositórios de AA nas principais universidades do estado de São Paulo, como forma de, consequentemente, prover AA aos artigos resultantes de pesquisas financiadas pela Fapesp, se tornaram uma opção tangível (Hey, 2017).

Em outubro de 2013, durante a sessão de abertura da 4ª Conferência Luso-Brasileira de AA (Confoa), foi lançado o Repositório da Produção Científica do Conselho de Reitores das Universidades Estaduais Paulistas (Cruesp)²⁴, contendo produções científicas das três principais universidades paulistas – a Universidade de São Paulo (USP), a Universidade Estadual Paulista (Unesp) e a Universidade Estadual de Campinas (Unicamp). No repositório estão disponíveis teses e dissertações, livros, artigos, trabalhos apresentados em reuniões científicas, entre outras tipologias documentais, em acesso livre para todos os interessados (Alisson, 2013).

Além disso, a Fapesp também vem articulando, com outras agências de fomento do Brasil, para motivar e expandir a iniciativa ao aconselhar que os resultados das pesquisas apoiados com investimentos públicos devam compor publicações em AA.

Em 2019, como forma de formalizar suas recomendações, a Fapesp, por meio da Portaria do Conselho Técnico-Administrativo (CTA) nº 01/2019, instituiu a Política para AA às Publicações Resultantes de Auxílios e Bolsas Fapesp²⁵. A política dispõe que os autores cujos artigos resultem de auxílios e bolsas financiadas pela fundação “deverão divulgá-los em periódicos que permitam o arquivamento de uma cópia do trabalho” em repositórios institucionais públicos, nos quais possam ser livremente acessados via *web*, descreve Marques (2019). Portanto, por meio dessa política, a Fapesp aconselha que os autores podem utilizar as diversas vias de publicação que priorizem o AA, desde que possam depositar uma cópia do artigo final em algum repositório público, com exceção daqueles que publicarem em algum periódico que faça parte da rede SciELO. De acordo com Mugnaini, Igami e Kryzanowski (2022), a Fapesp tem buscado maneiras de diminuir os custos relacionados às APCs dos artigos financiados, especialmente por meio de acordos com editoras que adotam o modelo híbrido de publicação.

24 Repositório da Produção Científica CRUESP. Disponível em: <http://cruesp.aguia.usp.br/>.

25 Política para AA às Publicações Resultantes de Auxílios e Bolsas Fapesp. Disponível em: <https://fapesp.br/12592/anexo-i-politica-para-acesso-aberto-as-publicacoes-resultantes-de-auxilios-e-bolsas-fapesp>.

Acompanhando tendências mundiais do movimento da OS – que também agrega o AA –, em dezembro de 2019, a Fapesp lançou a Rede de Repositórios de Dados Científicos do Estado de São Paulo²⁶, que envolve diversas instituições de pesquisas e universidades importantes. Por meio da plataforma é possível ter acesso aos dados gerados em pesquisas científicas, independentemente de sua publicação em artigos científicos, para atender à Política de Gestão de Dados da Fapesp²⁷.

Em outubro de 2021, por meio da Portaria CTA nº 37, a Fapesp reformulou a política de AA às publicações científicas, implementada em 2008. Tal reformulação estabelece um prazo máximo de 12 meses para disponibilização em AA das publicações oriundas de projetos financiados. A Fapesp espera que essa iniciativa aumente o acesso via internet, a visibilidade e os impactos científicos das pesquisas financiadas (Fundação de Amparo à Pesquisa do Estado de São Paulo, 2021).

Também vale ressaltar que a Fapesp foi a primeira agência de fomento a disponibilizar, por acesso livre, as informações referenciais de projetos apoiados por meio da Biblioteca Virtual da Fapesp (BV Fapesp)²⁸. O objetivo é tornar disponíveis as informações sobre as bolsas e auxílios, assim como os seus resultados, para serem facilmente acessados, nacional e internacionalmente (Freire, 2014). O estudo feito por Mugnaini, Igami e Kryzanowski (2020) demonstra que a Fapesp aparece em 30% das menções de financiamento de pesquisa dos artigos nacionais publicados, assim como a BV Fapesp possui, em sua base de dados, o mapeamento de 75% da produção científica apoiada pela fundação. Com isso, nota-se que a preocupação da Fapesp em participar, ao nível nacional e internacional, dos avanços do técnico-científico mundial acaba por agregar aspectos importantes ao estímulo, mapeamento e monitoramento do movimento do AA em âmbito nacional.

Hoje, a SciELO é um dos principais portais de AA do mundo. Evidencia-se, portanto, que as iniciativas em AA dessa magnitude desempenham um papel

26 Metabuscador de dados de pesquisa. Disponível em: <https://metabuscador.uspdigital.usp.br/>.

27 Gestão de Dados Fapesp. Disponível em: <https://fapesp.br/gestaodedados>.

28 Biblioteca Virtual da Fapesp. Disponível em: <https://bv.fappesp.br>.

importante para aumentar a visibilidade da ciência conduzida no Brasil e na América Latina.

Como descrevem Carvalho e demais autores (2017), o impulso para o AA cresce à medida que várias agências de fomento e instituições em todo o mundo implementam políticas de AA. No entanto, de acordo com Mugnaini, Igami e Kryzanowski (2022), devem ser observados os pontos das estratégias de internacionalização promovidas pelas agências de fomento que ainda reforçam o acesso restrito.

Outro ponto crítico para o sucesso das políticas de AA é a disponibilidade da infraestrutura necessária para implementá-las. A UE, por exemplo, tem se esforçado para promover o acesso livre das pesquisas financiadas com investimento público no âmbito do programa de pesquisa Horizon 2020. Em 2013, os Estados Unidos anunciaram que os artigos, resultados de pesquisas apoiadas com dinheiro público, devem ficar em AA em até um ano após sua publicação (Oliveira, 2017).

Embora o Brasil seja destaque quando se trata em AA, sendo um dos países mais ativos nessas frentes (Oliveira, 2017), nota-se que o papel das agências de fomento é fundamental no reconhecimento do movimento do AA em âmbito nacional, cujos resultados são muito concretos e motivam novas iniciativas nesse sentido. O desafio maior envolve os impactos institucionais da descentralização e das disparidades sociais e regionais. Portanto, as FAPs são fundamentais na articulação e na promoção dessas iniciativas e no alinhamento de políticas que motivem o AA.

A articulação entre as agências estaduais e federais deve centrar atividades de conscientização, estímulo, fortalecimento e avanço de políticas públicas em prol do AA no Brasil, “implicando, principalmente, a definição de agendas, propostas e ações dotadas de flexibilidades que permitam a operação conjunta de atores com diferentes características, conexões, habilidades e capacidades”, destacam Tonini, Santana e Scherrer (2020, p. 211).

O Brasil ainda segue na necessidade de investir na construção de repositórios de acesso livre e no estímulo para a publicação em periódicos científicos eletrônicos de AA. Historicamente, os órgãos de financiamento científico foram e são um dos principais influenciadores no *modus operandi* das comunidades científicas. Também são cruciais na articulação de ações

que visem à promoção do AA à literatura científica, que são de suma importância para a difusão do conhecimento científico. A filosofia subjacente do AA é que a pesquisa beneficia toda a humanidade e deve estar disponível para todos. Dessa forma, promover o AA ao conhecimento traz inúmeros benefícios para a sociedade em geral.

Desafios do monitoramento do AA

A implementação de políticas de promoção de publicações em AA desenvolveu uma demanda por monitoramento preciso do número absoluto e relativo de publicações em AA. No entanto, o AA é notoriamente difícil de medir e as análises frequentemente empregam técnicas de amostragem aleatória (Archambault *et al.*, 2014; Björk *et al.*, 2010). Todos os registros de publicação em uma determinada amostra são testados para determinar a proporção de textos completos que são publicações de AA.

Este método inevitavelmente introduz muitos erros de medição. De acordo com Van Leeuwen, Tatum e Wouter (2018), a implementação de novos sistemas de informação de pesquisa em universidades e institutos de pesquisa holandeses criou uma oportunidade para monitorar a parcela de publicações de AA em nível nacional por meio de esquemas de metadados coordenados e práticas de registro comuns. A avaliação da publicação em AA é complicada por uma diversidade crescente do que é considerado AA, pelas restrições de direitos autorais para quando uma publicação pode se tornar abertamente acessível e pela falta de identificação clara e consistente de publicações de AA em dados bibliográficos. Desde a publicação da BOAI em 2002, as definições de AA evoluíram para incluir uma variedade de formatos ou tipos de AA que aumentam a complexidade do acompanhamento do progresso em direção a um maior acesso à pesquisa publicada. Além das amplas categorias de vias dourada e verde de AA, podem existir várias versões de um manuscrito devido a variações nos acordos de licenciamento das editoras.

Esses acordos geralmente especificam como, quando e sob quais condições um manuscrito pode ser abertamente acessível na *web*. Por exemplo, um editor pode permitir a Via Verde de AA por meio do autoarquivamento

em um repositório institucional. No entanto, as restrições de direitos autorais dos editores diferem em vários estágios de desenvolvimento do manuscrito, atribuindo diferentes direitos de acesso a diferentes versões do texto. Tipos de versão comumente especificados incluem o manuscrito submetido – antes da revisão por pares –, o manuscrito aceito – revisado por pares, mas não formatado – e uma cópia exata do manuscrito publicado (Björk *et al.*, 2013). Isso cria a possibilidade de que a versão AA de um manuscrito seja substancialmente diferente da versão publicada. Em tais casos, não está claro se a versão AA foi validada por meio de medidas de controle de qualidade, como revisão por pares.

O acesso retardado/atrasado – Delayed Open Access – cria outra variação do AA. Após um determinado período de embargo, uma cópia da publicação pode ser autoarquivada ou o editor pode remover completamente as restrições de acesso ao *site* do periódico. Os períodos de embargo são geralmente especificados como um atraso de 6, 12, 18 ou 24 meses após a publicação, com 12 meses sendo o embargo mais comum (Laakso; Björk, 2013). No modo verde de AA, é deixado para os autores e para as instituições rastrear e gerenciar uma variedade de políticas de autoarquivamento, que por si só já pode ser barreira para AA (Davis; Connolly, 2007). Essa sobrecarga administrativa está amplamente ausente no caso de periódicos de assinatura que convertem artigos em AA após um atraso especificado – por exemplo, 12 meses. De acordo com Laakso e Björk (2013), os fatores de impacto de periódicos e artigos de periódicos de “acesso retardado/atrasado” são maiores que as médias comparáveis de periódicos por assinatura e periódicos de AA diretos – sem atraso.

Um tema comum em argumentos para o AA é que a publicação do AA aumenta o impacto da citação. Embora haja relatórios conflitantes em relação a essa “vantagem de citação de AA” – Open Access Citation Advantage (OACA) –, a atenção redobrada a esse problema aumentou o entendimento sobre o comportamento de citação de maneira mais geral. Alguns estudos bibliométricos afirmam que a publicação de AA resulta em um aumento significativo nas citações (Van Leeuwen; Tatum; Wouter; 2018).

Há estudos que ressaltam que o AA favorece o aumento do impacto, ou seja, artigos de AA são mais citados do que aqueles com acesso restrito

(Hitchcock, 2013). Dentre os principais estudos, destaca-se o de Lawrence (2001), que foi o primeiro a publicar dados reconhecendo a tendência de que a publicação *on-line* pode aumentar o impacto. Harnad e Brody (2004) confirmam ainda que essa tendência se aplica aos artigos de AA pelo trabalho desenvolvido no âmbito do Open Citation Project²⁹ baseado no ArXiv. A importância desse estudo foi destacada por Peter Suber no “Open Access News”, quando afirmou que se tratava de um artigo importante, pois era o primeiro grande estudo desde o famoso artigo de Lawrence que documentou a proposição de que o AA aumenta o impacto. É também o primeiro a ir além de Lawrence em escopo e método, a fim de responder às dúvidas levantadas sobre sua tese. Ao confirmar que o AA aumenta o impacto, fornece aos autores as melhores razões para publicarem seus trabalhos em fontes abertas.

Colaborações mais amplas surgiram para estender esses achados que ressaltam que o AA se tornou viável devido à mudança para a publicação e divulgação *on-line* (Brody *et al.*, 2004). Uma nova medida que se tornou possível com a publicação *on-line* é o número de *downloads*, abrindo uma nova linha de investigação. Brody e demais autores (2004) foram proeminentes em mostrar que existe uma correlação entre uma quantidade maior de *downloads* e maior impacto de citação.

Isso poderia funcionar como um complemento às métricas de mensuração tradicionais, pois por meio da análise da quantidade de *downloads* poderia se prever mais cedo quais as publicações terão maior impacto, sem a necessidade de se esperar anos para que as citações se materializem em outras publicações.

McCabe e Snyder (2013) descobriram que passar do acesso pago para o AA aumentou as citações, em média, em 8% na amostra, mas o efeito variou de acordo com a qualidade do conteúdo. O AA aumentou as citações para o melhor conteúdo – periódicos bem classificados ou artigos nos quintis superiores de citações dentro de um volume –, mas reduziu as citações para conteúdo de qualidade inferior, ou seja, colocar o periódico em uma ampla plataforma de AA pode aumentar a competição entre os artigos pela atenção dos leitores.

29 Open Citations Homepage. Disponível em: <https://opencitations.net/>.

Nesses estudos, o tamanho da vantagem nessa relação entre AA e citações varia amplamente com base em uma variedade de questões, como diferenças entre áreas do conhecimento, abordagens metodológicas, variação na forma como o AA é definido e a dificuldade em determinar quando um artigo se torna abertamente acessível (Swan, [2010]). Além disso, os principais fatores que influenciam a frequência de citação estão ou podem estar relacionados à exposição precoce a versões preliminares de um manuscrito (Moed, 2007), à disponibilidade em múltiplas fontes de informação (Xia; Myers; Wilhoite, 2010) e à proximidade física dos pesquisadores (Lee *et al.*, 2010).

Em um estudo mais recente, Archambault e demais autores (2014) mostraram variação no acúmulo de citações associadas a diferentes modos de AA. Os autores encontraram uma vantagem de citação mais proeminente associada ao modo de autoarquivamento de AA verde e uma desvantagem de citação associada a periódicos AA pela Via Dourada.

Diante desse cenário, vale ressaltar a importância do monitoramento tanto das publicações em AA como também do impacto e da relevância nas comunidades científicas, inclusive para elaboração de políticas para implementação e avaliação. No entanto, diante da diversidade e da complexidade dos fatores envolvidos para a efetividade do AA, as comunidades necessitam se debruçar em estudos que contemplem medidas de apoio, não somente relacionadas ao fomento científico, mas também à implementação de infraestrutura necessária para elaboração de indicadores de mensuração.

Experiências internacionais de monitoramento ao AA

Conforme destacado pelo *Open Access Monitoring: Guidelines and Recommendations for Research Organisations and Funders*, publicado em 2021 pela Science Europe, o monitoramento do AA permite uma visão mais profunda das tendências de publicação, podendo contribuir na elaboração de estratégias futuras em níveis institucionais e nacionais, fornecendo orientação para o desenvolvimento e para a revisão de políticas, ajudando a avaliar os efeitos dos mecanismos de financiamento, que é crucial para negociar acordos transformadores com editores de assinaturas tradicionais. Além disso,

contribui para uma base factual para as decisões de apoio ao movimento do AA (Science Europe, 2021). Vale destacar algumas iniciativas de monitoramento mencionadas no relatório.

Em 2017, o Universities UK Open Access Coordination Group³⁰ (Grupo de Coordenação de AA das Universidades do Reino Unido) com múltiplas partes interessadas publicou seu segundo relatório analisando o progresso do Reino Unido em relação ao AA. Usando dados da Scopus e de várias outras fontes de dados, o relatório considerou tendências em vários aspectos de publicação de AA, como o *status* da publicação, o licenciamento e os custos.

A fim de avaliar a conformidade dos pesquisadores com seus requisitos de AA, o Austrian Science Fund (FWF)³¹ monitora a taxa de AA de publicações revisadas por pares em projetos financiados pelo FWF. A base para os cálculos é o *status* de AA das publicações no momento dos relatórios finais do projeto.

Desde 2018, o Danish Service for Science and Higher Education³² (Serviço Dinamarquês para Ciência e Ensino Superior) produz anualmente o Danish Open Access Indicator – Indicador de AA Dinamarquês. Os resultados são interpretados e apresentados no âmbito da estratégia nacional de desenvolvimento científico e tecnológico. Para o indicador, um fluxo de trabalho padronizado baseia-se em dados do Current Research Information Systems (Cris) das próprias universidades dinamarquesas, enriquecido com metadados do Directory of Open Access Journals e SHERPA/RoMEOB³³, que é um serviço que fornece informações sobre as políticas de AA dos editores.

O French Open Science Monitor³⁴ foi desenvolvido no contexto do Plano Nacional Francês para a Ciência Aberta. A ferramenta é baseada em

30 Monitoring the transition to open access: December 2017. Disponível em: <https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Pages/monitoring-transition-open-access-2017.aspx>.

31 FWF Der Wissenschaftsfonds. Disponível em: <https://www.fwf.ac.at/de/ueber-den-fwf/publikationen/publication-types/43/>.

32 Danish Open Access Indicator. Disponível em: <https://www.oaindikator.dk/en/>.

33 About Sherpa Romeo. Disponível em: <https://v2.sherpa.ac.uk/romeo/about.html>.

34 Ouvrir La Science. French Open Science Monitor. Disponível em: <https://www.ouvrirlas-cience.fr/the-french-open-science-monitor/>.

dados disponíveis abertamente e recorre a sistemas de grande escala, como Unpaywall, HALD – o principal repositório aberto na França, parte do Centro Nacional Francês de Pesquisa Científica (CNRS) –, ORCID e IDRef (referencial para ensino superior francês e pesquisa). Ele usa técnicas baseadas em regras e no aprendizado de máquina para enriquecer os metadados das publicações.

O German Open Access Monitor³⁵, mede a produção de publicações de instituições de pesquisa alemãs que recebem financiamento público, ao mesmo tempo em que informa sobre o *status* do AA. A maneira como interpretam e apresentam os dados é considerada amigável, pois facilita a reutilização dos dados. Esses dados de publicação são obtidos por meio do Dimensions e WoS, a partir de 2021 dados da Scopus devem ser adicionados. Os metadados sobre o *status* de AA são coletados do Unpaywall.

Outra iniciativa para fornecer *insights* sobre o *status* de AA de publicações de pesquisa em nível mundial, a Australian Curtin Open Knowledge Initiative³⁶ (COKI) baseia-se em dados de fontes públicas em todo o mundo, dando uma visão geral das estatísticas de AA para muitos países. Os dados da publicação são coletados do Crossref e do Microsoft Academic. As informações do “Funder Registry” da Crossref³⁷ e do Global Research Identifier Database³⁸ (GRID) são usadas para vincular publicações a financiadores e instituições de pesquisa.

Desde 2019, o CWTS Leiden Ranking³⁹ inclui o número e a proporção de publicações de AA como uma métrica para avaliar o desempenho científico de mil universidades em todo o mundo. O CWTS usa dados de publicação

35 About Open Access Monitor. Disponível em: <https://open-access-monitor.de/>.

36 COKI Open Access Dashboard. Disponível em: <http://openknowledge.community/dashboards/coki-open-access-dashboard/>.

37 Crossref. Funder Registry. Disponível em: <https://www.crossref.org/services/funder-registry/>.

38 Global Research Identifier Database (GRID). Disponível em: <https://www.grid.ac/>.

39 CWTS Leiden Ranking. Disponível em: <https://www.leidenranking.com/>.

da WoS e metadados de AA da Unpaywall. Seus indicadores diferenciam diferentes tipos de AA usando cores⁴⁰.

Complexidades do monitoramento ao AA no Brasil

O Brasil tem sido internacionalmente reconhecido por suas estratégias bem-sucedidas de AA, por influenciar outros países para o movimento (Weitzel, 2019). As principais iniciativas para a formação de agendas que estão relacionadas à comunicação científica aberta são coordenadas pela SciELO, pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), pela Fapesp⁴¹, pela USP, pela Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e pela Fundação Oswaldo Cruz (Fiocruz)⁴²⁻⁴³. Embora o Brasil seja o berço de práticas pioneiras e bem-sucedidas de AA, ainda é um desafio modificar o *status quo* do reconhecimento científico definido há séculos, em que é necessária a publicação em periódicos de alto fator de impacto.

De acordo com Xavier (2019, p. 498), durante o início da década de 2010, o número de repositórios digitais duplicou, assim como “o desenvolvimento da via verde no Brasil teve seu maior pico de crescimento”. O mesmo estudo demonstra que, entre 2016 e 2019, o ritmo de crescimento ocorreu moderadamente e destaca que:

com maior volume e menos períodos de estagnação do que a Via Verde, o desenvolvimento da Via Dourada no Brasil no mesmo período se deu de modo diferente. Seu maior pico de crescimento ocorreu somente na segunda metade da década de 2010. Até metade de 2014, o registro de periódicos científicos cresceu 79%. Embora não seja uma marca ruim, o período à frente, de 2015 até o início de 2019, superou amplamente essa marca, registrando aumento de 167% (Xavier, 2019, p. 498).

40 CWTS Leiden Ranking. Indicators. Disponível em: <https://www.leidenranking.com/information/indicators>.

41 Fapesp. Boas práticas científicas. Disponível em: <https://fapesp.br/boaspraticas/>.

42 Veja mais em: Santos e demais autores. (2014).

43 No desenvolvimento e implantação do Repositório Institucional Arca Fiocruz. Disponível em: <https://portal.fiocruz.br/repositorio-institucional-arca>.

Por sua vez, a pesquisa de Weitzel (2019) possuiu o objetivo de mapear os repositórios institucionais brasileiros até maio de 2017, com objetivo de observar o atual estado das estratégias de AA. A autora destaca que no Brasil o IBICT foi o principal articulador do AA, sobretudo em Via Verde, e envolveu diversas ações, tanto em prover infraestruturas necessárias, quanto ao apoio ao AA no país, de modo a acompanhar os acontecimentos de AA mundiais. O IBICT também é o aliado na promoção de um dos principais eventos sobre a temática, a Confoa⁴⁴, que reúne pesquisadores brasileiros e portugueses em debates sobre as questões a respeito do AA e da ciência aberta.

Nos últimos anos, Santos e Lima (2015) descrevem que o crescimento de políticas públicas de AA de caráter mais mandatário e a constituição de repositórios institucionais foram bastante significativos no Brasil, mas ainda ocorrem lentamente. Editais da Finep em convênio com o IBICT, lançados a partir de 2009, buscam o fortalecimento do AA por meio da distribuição de *kits* tecnológicos para implementação de publicações digitais e de repositórios institucionais com uso da plataforma DSPACE (Weitzel, 2019).

Embora o crescimento do AA no país tenha sido perceptível e esteja na frente de países como os Estados Unidos quando se trata da Via Dourada, o Brasil ainda não consegue se equiparar a esses países quando se trata da Via Verde (Weitzel, 2019). Isso significa que ainda são necessários maiores investimentos e esforços para aumentar a quantidade de repositórios institucionais e articulá-los conjuntamente a favor de objetivos em comum. Por outro lado, também é necessário lidar com os complexos processos de gestão e manutenção dos repositórios, que se tornam cada vez mais complexos e robustos (Weitzel, 2019).

Santos e Lima (2015) ressaltam que o AA ainda precisa ser muito debatido para ser incorporado estrategicamente pelas instituições de ensino e pesquisa brasileiras, pois somente dessa maneira será possível criar alguns parâmetros para o seu monitoramento. Percebe-se que a valorização do AA ocorreu concomitantemente à valorização da educação, da pesquisa e da ciência brasileiras. Os números tendem a melhorar a partir do aumento da

44 Sobre a Confoa. Disponível em: <https://conferencias.rcaap.pt/confoa/index/about>.

qualidade dos repositórios institucionais, da visibilidade e dos impactos da produção científica brasileira, para que sejam valorizadas por sua relevância na cobertura de pesquisas orientadas para a resolução de problemas locais e nacionais, estruturalmente diferentes da ciência predominante (*mainstream*). Quanto mais o movimento evolui, torna-se necessário envolver as instituições brasileiras de diversas instâncias públicas. Para tanto, ainda são necessárias diretrizes, com a articulação de ações regionais e integradas para o fomento de políticas e programas nacionais de AA, além de ações que visem à contínua promoção de boas práticas, treinamento e investimento em infraestrutura.

Isso demonstra que diversas pesquisas devem ser feitas, nas quais são exigidas análises mais profundas para monitorar as tendências de AA e os demais fatores que ajudam a avaliar o crescimento do movimento em âmbito nacional.

Considerações finais

Um dos principais objetivos das políticas de AA é que os artigos científicos sejam não apenas mais acessíveis, mas também mais amplamente lidos e citados. Uma forma de resolver parte das questões de monitoramento seria o monitoramento dos *downloads* dessas publicações, embora se entenda que o fato de ter feito *download* não signifique garantia do uso, mas serve como um indicativo de sua utilização. Todavia, não é tarefa fácil a realização desse tipo de levantamento, já que as publicações podem ser obtidas de diferentes fontes de informação, sejam em periódicos, *sites* pessoais, repositórios institucionais ou pelas mídias sociais.

Vale depreender que os próprios repositórios institucionais diferem em suas características e gestão. Por isso, é necessário o monitoramento de diferentes plataformas e com uso de tecnologias específicas que permitam a varredura em grande escala. Para tal, é preciso reforçar a conscientização e o alinhamento dos propósitos que beneficiem as orientações acerca das vantagens do AA.

Percebe-se nitidamente que existem muitos desafios para o monitoramento de AA. Dentre eles, destaca-se a necessidade de metadados descritivos eficientes, pois diferentes fontes de informação deverão ser agregadas

e conciliadas; a identificação dos autores para que não haja equívocos em relação à afiliação e às fontes do financiamento, quando houver. Assim, os identificadores persistentes se mostram importantes para a efetividade da análise das informações, sendo também um limitador, já que são recursos recentes em comparação à imensidão de publicações científicas que não receberam o DOI, por exemplo.

Outro desafio se dá diante da diversidade documental disponível em repositórios nacionais, que variam constantemente dependendo dos propósitos de criação dessas plataformas. Em geral, os repositórios brasileiros possuem coleções mescladas, com abordagens flexíveis que, de acordo com Leite (2009), visam ampliar o acesso de conteúdos de comunicação e dos resultados da pesquisa científica, além da produção científica avaliada por pares. Se por um lado muito se discute sobre quais tipologias documentais aderem aos objetos do movimento do AA, dos quais reforçam a abordagem do AA verde, por outro, essa visão se torna efetiva quando se trata da OS que, de forma ampla, também abarca o AA.

Cabe mencionar que o fomento de infraestrutura de AA no Brasil, sobretudo da distribuição de repositórios por região, ainda ocorre principalmente na Região Sudeste, sendo necessárias mais ações para resolução de disparidades científico-tecnológicas regionais, desafio também para as questões de descentralização das políticas públicas. Além disso, muitos dos investimentos feitos na criação de repositórios ainda são, de contrapartida, das próprias instituições, denotando que o apoio em infraestruturas necessita ser observado adequadamente pelas esferas públicas (Weitzel, 2019).

A questão central que se coloca é a necessidade urgente de aperfeiçoamento dos instrumentos contemporâneos de acompanhamento e avaliação da produção em CT&I, considerando seu impacto na sociedade.

Embora tenha trazido aspectos importantes para o debate do AA, este trabalho faz sua contribuição ao trazer um robusto referencial teórico e bibliográfico para continuidade das pesquisas acerca da temática.

Referências

- AGUILLO, I. F. *et al.* Indicators for a webometric ranking of open access repositories. *Scientometrics*, [Budapest], v. 82, p. 477-486, 2010. DOI 10.1007/s11192-010-0183-y. Disponível em: <https://link-springer-com.ez10.periodicos.capes.gov.br/article/10.1007/s11192-010-0183-y>. Acesso em: 25 jun. 2021.
- ALISSON, E. USP, Unesp e Unicamp disponibilizam produção científica na internet. *Agência FAPESP*, São Paulo, 10 out. 2013. Disponível em: <https://agencia.fapesp.br/usp-unesp-e-unicamp-disponibilizam-producao-cientifica-na-internet/18026/>. Acesso em: 12 jan. 2021.
- ALVES, M. C. A SBPC e as fundações de amparo à pesquisa. *Ciência e Cultura*, São Paulo, v. 70, n. 4, p. 8-10, out./dez. 2018. DOI 10.21800/2317-66602018000400003. Disponível em: <http://cienciaecultura.bvs.br/pdf/cic/v70n4/v70n4a03.pdf>. Acesso em: 2 jan. 2021.
- ANDRADE, R. de M.; MURIEL-TORRADO, E. Declarações de Acesso Aberto e a Lei de Direitos Autorais brasileira. *RECIIS - Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, Rio de Janeiro, v. 11, p. 1-5, nov. 2017. DOI 10.29397/reciis.v11i0.1374. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/1374/pdf1374>. Acesso em: 25 jan. 2020.
- ARCHAMBAULT, É. *et al.* Proportion of open access papers published in peer-reviewed journals at the European and world levels: 1996-2013. [S. l.]: European Commission, 2014. Disponível em: https://science-metrix.com/sites/default/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf. Acesso em: 25 jan. 2020.
- AVENTURIER, P. Unpaywall, uma extensão Chrome para achar a versão dos artigos em acesso aberto. [S. l.], 2 maio 2017. *Blog do Pascal Aventurier*. Disponível em: <https://publicient.hypotheses.org/1715>. Acesso em: 25 jan. 2020.
- BAILEY JUNIOR, C. W. *Open access bibliography: liberating scholarly literature with e-prints and open access journals*. Washington: Association of Research Libraries, 2005. Disponível em: <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1154&context=scholcom>. Acesso em: 27 jul. 2021.

BJÖRK, B.-C. *et al.* Anatomy of green open access. *Journal of the Association for Information Science and Technology*, New York, v. 65, n. 2, p. 237-250, 2013. DOI 10.1002/asi.22963. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/asi.22963>. Acesso em: 12 abr. 2020.

BJÖRK, B.-C. *et al.* Open access to the scientific journal literature: situation 2009. *PLoS ONE*, California, v. 5, n. 6, e11273, 2010. DOI 10.1371/journal.pone.0011273. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011273>. Acesso em: 12 jun. 2021.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. *CAPES e CNPq incentivam publicações científicas*. Brasília, DF, 24 set. 2007. Disponível em: <https://www.gov.br/capes/pt-br/assuntos/noticias/blank-76396233>. Acesso em: 22 jan. 2021.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. *Ministro da Educação lança Portal de Acesso Livre da Capes*. Brasília, DF, 31 out. 2022. Disponível em: <https://www.gov.br/capes/pt-br/assuntos/noticias/blank-27742552>. Acesso em: 7 out. 2024.

BRODY, T. *et al.* The effect of Open Access on Citation Impact. In: NATIONAL POLICIES ON OPEN ACCESS (OA) PROVISION FOR UNIVERSITY RESEARCH OUTPUT: AN INTERNATIONAL MEETING, 2004, Southampton. *Proceedings* [...]. Southampton: Southampton University, 2004. Disponível em: <https://eprints.soton.ac.uk/259941/1/OATAnew.pdf>. Acesso em: 12 jun. 2021.

CARVALHO, J. *et al.* Monitoring a national open access funder mandate. *Procedia Computer Science*, [s. l.], v. 106, p. 283-290, 2017. DOI 10.1016/j.procs.2017.03.027. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050917302958>. Acesso em: 25 jun. 2020.

CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS. *Descentralização do fomento à ciência, tecnologia e inovação no Brasil*. Brasília, DF: CGEE, 2010. Disponível em: <https://ses.sp.bvs.br/wp-content/uploads/2017/07/Descentralizacao-do-fomento-cti-no-brasil.pdf>. Acesso em: 15 mar. 2021.

COSTA, M. P. da; LEITE, F. C. L. Open access in the world and Latin America: a review since the Budapest Open Access Initiative. *Transinformação*, Campinas, v. 28, n. 1, p. 33-45, jan./abr. 2016. DOI 10.1590/2318-08892016002800003. Disponível em: <https://www.scielo>.

br/j/tinf/a/3wTyrH3gwFcxSXrvs5B6NnS/?lang=en&format=pdf. Acesso em: 25 jun. 2020.

COSTA, S. M. S. Filosofia aberta, modelos de negócios e agências de fomento: elementos essenciais a uma discussão sobre o acesso aberto à informação científica. *Ciência da Informação*, Brasília, DF, v. 35, n. 2, p. 39-50, maio/ago. 2006. DOI 10.1590/S0100-19652006000200005. Disponível em: <https://www.scielo.br/j/ci/a/XsgXnnC7xWHNR7gXrP9Hw3M/?format=pdf&lang=pt>. Acesso em: 25 jun. 2020.

CRUZ, C. H. de B. An accelerator for science in Latin America: FAPESP. *Elephant in the Lab*, [s. l.], 26 nov. 2019. DOI 10.5281/zenodo.3552277. Disponível em: <https://elephantinthelab.org/an-accelerator-for-science-in-latin-america-fapesp/>. Acesso em: 12 jan. 2021.

DAVID, P. A. The historical origins of Open Science: an essay on patronage, reputation and common agency contracting in the scientific revolution. *Capitalism and Society*, [s. l.], v. 3, n. 2, 2008.

DAVIS, P. M.; CONNOLLY, M. J. L. Institutional repositories: evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine*, [s. l.], v. 13, n. 3/4, 2007. DOI 10.1045/march2007-davis. Disponível em: <https://www.dlib.org/dlib/march07/davis/03davis.html>. Acesso em: 25 jun. 2020.

DUNN, A. G.; COIERA, E.; MANDL, K. D. Is biblioleaks inevitable? *Journal of Medical Internet Research*, [s. l.], v. 16, n. 4, e112, 2014. DOI 10.2196/jmir.3331. Disponível em: <https://www.jmir.org/2014/4/e112/>. Acesso em: 25 jun. 2020.

ELSE, H. Funders flesh out details of Europe's bold open-access plan. *Nature*, [s. l.], 27 Nov. 2018b. DOI 10.1038/d41586-018-07557-w. Disponível em: <https://www.nature.com/articles/d41586-018-07557-w>. Acesso em: 25 jun. 2020.

ELSE, H. Radical open-access plan could spell end to journal subscriptions. *Nature*, [s. l.], v. 561, p. 17-18, 2018a. Disponível em: <https://www.nature.com/articles/d41586-018-06178-7>. Acesso em: 25 jun. 2020.

ELSE, H. Web of Science Owner Buys Tool that Offers One-click Access to Journal Articles. *Nature*, [s. l.], 10 Apr. 2018c. DOI 10.1038/d41586-018-

04414-8. Disponível em: <https://www.nature.com/articles/d41586-018-04414-8>. Acesso em: 25 jun. 2020.

FAUSTO, S. Evolução do Acesso Aberto: breve histórico. *SciELO em Perspectiva*, [São Paulo], 21 out. 2013. Disponível em: <https://blog.scielo.org/blog/2013/10/21/evolucao-do-acesso-aberto-breve-historico/#.YQxRKOhKjIU>. Acesso em: 12 jan. 2021.

FREIRE, D. Biblioteca Virtual da FAPESP é apresentada como exemplo de acesso aberto. *Agência FAPESP*, São Paulo, 15 out. 2014. Disponível em: <https://agencia.fapesp.br/biblioteca-virtual-da-fapesp-e-apresentada-como-exemplo-de-acesso-aberto/20033/>. Acesso em: 12 jan. 2021.

FREIRE, J. D. *CNPq e o acesso aberto à informação científica*. 2011. Tese (Doutorado em Ciência da Informação) – Faculdade de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, DF, 2011.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. FAPESP reformula sua política de acesso aberto a publicações. *Agência FAPESP*, São Paulo, 10 nov. 2021. Disponível em: <https://agencia.fapesp.br/fapesp-reformula-sua-politica-de-acesso-aberto-a-publicacoes/37271/>. Acesso em: 16 nov. 2021.

GADHVI, M.; SRIVASTAV, S.; SHARMA, R. K. Access to scientific literature: legitimate channels. *Indian Journal of Physiology and Pharmacology*, New Delhi, v. 64, n. 2, p. 155-157. Apr./June 2020. DOI 10.25259/IJPP_107_. Disponível em: <https://ijpp.com/view-pdf/?article=0fd912e9d3ed8e2d5b6c8e91a35bc175gq86FAp7qYU=>. Acesso em: 25 jun. 2020.

GORRAIZ, J.; WIELAND, M. Multi-authored publications: their influence in the distribution of the financing costs in world licenses. *Research Evaluation*, Oxford, v. 18, n. 3, p. 215-220, 2009.

GORRAIZ, J.; WIELAND, M.; GUMPENBERGER, C. Bibliometric practices and activities at the University of Vienna. *Library Management*, [s.l.] v. 33, n. 3, p. 174-183, 2012.

HARNAD, S. Mandates and metrics: how open repositories enable universities to manage, measure and maximise their research assets. Southampton: University of Southampton Institutional Repository

Preprint, 2007. 99 slides. Disponível em: <https://eprints.soton.ac.uk/264990/>. Acesso em: 25 jun. 2020.

HARNAD, S.; BRODY, T. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine*, [s. l.], v. 10, n. 6, jun. 2004. DOI 10.1045/june2004-harnad. Disponível em: <http://www.dlib.org/dlib/june04/harnad/06harnad.html>. Acesso em: 25 jun. 2020.

HARNAD, S. *et al.* The access/impact problem and the green and gold roads to open access: an update. *Serials Review*, Michigan, v. 34, n. 1, p. 36-40, 2008. DOI 10.1016/j.serrev.2007.12.005. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0098791307001530>. Acesso em: 25 jun. 2020.

HEY, T. A global view of Open Access (2): the perspective from Brazil and the SciELO open access portal. *TonyHey.net*, [s. l.], 10 Mar. 2017. Disponível em: <https://tonyheynet.com/2017/03/10/a-global-view-of-open-access-2-the-perspective-from-brazil-and-the-scielo-open-access-portal/>. Acesso em: 12 jan. 2021.

HITCHCOCK, S. (ed.). *The effect of open access and downloads ('hits') on citation impact: a bibliography of studies*. [S. l.], 2013. Disponível em: <https://eprints.soton.ac.uk/354006/1/oacitation-biblio-snapshot0613.html>. Acesso em: 25 jun. 2020.

HOOK, D. *Open Access surpasses subscription publication globally for the first time*. *Dimensions*, [s. l.], 24 Feb. 2021. Disponível em: <https://www.dimensions.ai/blog/open-access-surpasses-subscription-publication-globally-for-the-first-time/>. Acesso em: 26 fev. 2021.

KURAMOTO, H. Acesso livre: caminho para maximizar a visibilidade da pesquisa. *RAC - Revista de Administração Contemporânea*, Curitiba, v. 12, n. 3, p. 861-872, jul./set. 2008. DOI 10.1590/S1415-65552008000300013. Disponível em: <https://rac.anpad.org.br/index.php/rac/article/view/625/622>. Acesso em: 25 jun. 2020.

LAAKSO, M.; BJÖRK, B. C. Delayed open access: An overlooked high-impact category of openly available scientific literature. *Journal of the American Society for Information Science and Technology*, New York, v. 6, n. 7, p. 1323-1329, 2013.

LAAKSO, M. *et al.* The Development of Open Access Journal Publishing de 1993 a 2009. *PLoS ONE*, California, v. 6, n. 6, e20961, 2011. DOI 10.1371/journal.pone.0020961. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0020961>. Acesso em: 25 jun. 2020.

LAWRENCE, S. Free online availability substantially increases a paper's impact. *Nature*, [s. l.], v. 411, p. 521, 31 May 2001. DOI 10.1038/35079151. Disponível em: <https://www.nature.com/articles/35079151>. Acesso em: 25 jun. 2020.

LEE, K. *et al.* Does collocation inform the impact of collaboration? *PLoS ONE*, California, v. 5, n. 12, e14279, 15 Dec. 2010. DOI 10.1371/journal.pone.0014279. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0014279>. Acesso em: 25 jun. 2020.

LEITE, F. C. L. *Como gerenciar e ampliar a visibilidade da informação científica brasileira*: repositórios institucionais de acesso aberto. Brasília, DF: IBICT, 2009. Disponível em: <https://livroaberto.ibict.br/bitstream/1/775/4/Como%20gerenciar%20e%20ampliar%20a%20visibilidade%20da%20informa%c3%a7%c3%a3o%20cient%c3%adfica%20brasileira.pdf>. Acesso em: 18 jun. 2020.

LETA, J.; COSTA, E. H. S.; MENA-CHALCO, J. P. Artigos em Periódicos de Acesso Aberto: um Estudo com Pesquisadores Bolsistas de Produtividade do CNPq. *RECIIS - Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, Rio de Janeiro, v. 11, p. 1-6, nov. 2017. DOI 10.29397/reciis.v11i0.1408. Disponível em: <https://www.recis.icict.fiocruz.br/index.php/reciis/article/view/1408/pdf1408>. Acesso em: 12 jan. 2021.

MACCALLUM, C. J. When is open access not open access? *PLoS Biology*, [s. l.], v. 5, n. 10, p. 2095-2097, 2007. DOI 10.1371/journal.pbio.0050285. Disponível em: <https://journals.plos.org/plosbiology/article/file?id=10.1371/journal.pbio.0050285&type=printable>. Acesso em: 12 maio 2021.

MARQUES, F. FAPESP lança política para acesso aberto a publicações científicas. *Revista Pesquisa FAPESP*, São Paulo, 14 mar. 2019. Disponível em: <https://revistapesquisa.fapesp.br/fapesp-lanca-politica-para-acesso-aberto/>. Acesso em: 25 jan. 2021.

MARQUES, F. O futuro do acesso aberto: com oferta gratuita dos papers da União Europeia a partir de 2020, publicações científicas procuram novos modelos. *Revista Pesquisa FAPESP*, São Paulo, v. 245, jul. 2016. Disponível em: <https://revistapesquisa.fapesp.br/o-futuro-do-acesso-aberto/>. Acesso em: 12 abr. 2021.

MARTÍN-MARTÍN, A. *et al.* Evidence of open access of scientific publications in Google Scholar: a large-scale analysis. *Journal of Informetrics*, [s. l.], v. 12, n. 3, p. 819-841, Aug. 2018. DOI 10.1016/j.joi.2018.06.012. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1751157718301457>. Acesso em: 25 jun. 2020.

MCCABE, M. J.; SNYDER, C. M. The rich get richer and the poor get poorer: the effect of open access on cites to science journals across the quality spectrum. *SSRN - Social Science Research Network*, [s.l.] 25 maio 2013.

MELO, J. H. N. de; CAFÉ, L. C. Revistas de acesso aberto no portal de periódicos da CAPES. *Cadernos BAD*, Lisboa, n. 1, p. 309-319, 2018. Disponível em: <https://publicacoes.bad.pt/revistas/index.php/cadernos/article/view/1920/pdf>. Acesso em: 2 jun. 2021.

MENEZES, E. T. de. Verbete FAPs (Fundações de Amparo à Pesquisa). *Dicionário Interativo da Educação Brasileira*. São Paulo: Midiamix, 2001. Disponível em: <https://www.educabrasil.com.br/faps-fundacoes-de-amparo-a-pesquisa/>. Acesso em: 2 jan. 2021.

MOEDAS, C. Introduction. In: EUROPEAN COMMISSION. *Open innovation, open science, open to the world: a vision for Europe*. Luxembourg: European Commission, 2015. p. 6-7. Disponível em: http://publications.europa.eu/resource/ellar/3213b335-1cbc-11e6-ba9a-01aa75ed71a1.0001.02/DOC_2. Acesso em: 25 jun. 2020.

MUGNAINI, R.; IGAMI, M. P. Z.; KRYZANOWSKI, R. F. Acesso aberto e financiamento da pesquisa no Brasil: características e tendências da produção científica. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 27, p. 1-26, 2022. DOI 10.5007/1518-2924.2022.e. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/78818/48584>. Acesso em: 7 out. 2024.

MUGNAINI, R.; IGAMI, M. P. Z.; KRYZANOWSKI, R. F. Indexação e financiamento da pesquisa no Brasil: relações cientométricas. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 7.,

2020, Salvador. *Anais [...]*. Salvador: UFBA, 2020. Disponível em: <https://cip.brapci.inf.br//download/148932>. Acesso em: 20 ago. 2021.

OLIVEIRA, G. A. de. Políticas públicas e financiamento guiam a orientação de periódicos para acesso aberto ou fechado. *ComCiência: Revista Eletrônica de Jornalismo Científico*, [s. l.], 9 set. 2017. Disponível em: <https://www.comciencia.br/politicas-publicas-e-financiamento-guiam-a-orientacao-de-periodicos-para-acesso-aberto-ou-fechado/>. Acesso em: 12 jan. 2021.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA. *Open access infrastructure*. Paris: UNESCO, 2015. v. 2. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000232204?posInSet=1&queryId=4e826988-8c02-4be7-b76d-afd271dae39e>. Acesso em: 15 mar. 2021.

ORTELLADO, P. As políticas nacionais de acesso à informação científica. *Liinc em Revista*, Rio de Janeiro, v. 4, n. 2, p. 186-195, 2008. DOI 10.18617/liinc.v4i2.268. Disponível em: <https://revista.ibict.br/liinc/article/view/3164/2830>. Acesso em: 25 jun. 2020.

PARKER, A. L.; SANTOS, S. Ciência aberta e o novo modus operandi de comunicar pesquisa: parte I. *SciELO em Perspectiva*, [São Paulo], 1 ago. 2019. Disponível em: <https://blog.scielo.org/blog/2019/08/01/ciencia-aberta-e-o-novo-modus-operandi-de-comunicar-pesquisa-parte-i/#.YFT1pq9KjIV>. Acesso em: 25 jan. 2021.

PIWOWAR, H. *et al.* The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, [s. l.], v. 6, e4375, p. 1-23, 2018. DOI 10.7717/peerj.4375. Disponível em: <https://peerj.com/articles/4375/>. Acesso em: 25 jun. 2020.

REZENDE, L. V. R.; ABADAL, E. Estado da arte dos marcos regulatórios brasileiros rumo à Ciência Aberta. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis. v. 25, p. 1-25, 2020. DOI 10.5007/1518-2924.2020.e71370. Disponível em: <https://www.redalyc.org/journal/147/14763386043/html/>. Acesso em: 12 fev. 2021.

ROBINSON-GARCIA, N.; COSTAS, R.; Van LEEUWEN, T. N. 2020. Open Access uptake by universities worldwide. *PeerJ*, [s. l.], v. 8, e9410, p. 2-20, 8 jul. 2020. DOI doi.org/10.7717/peerj.9410. Disponível em: <https://peerj.com/articles/9410/>. Acesso em: 25 jul. 2020.

SANTOS, P. X. dos; LIMA, N. T. Acesso aberto: uma nova possibilidade de monitorar e avaliar o fluxo e o impacto da ciência. *PontodeAcesso*, Salvador, v. 9, n. 3, p. 149-163, 2015. DOI 10.9771/1981-6766rpa.v9i3.15166. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/15166/10331>. Acesso em: 15 fev. 2021.

SANTOS, P. X. dos *et al.* Política de acesso aberto ao conhecimento: análise da experiência da Fundação Oswaldo Cruz/Fiocruz. *RECIIS - Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, Rio de Janeiro, v. 8, n. 2, p. 210-226, 2014. DOI 10.3395/reciis.v8i2.633. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/633/1273>. Acesso em: 25 jun. 2020.

SCIENCE EUROPE. *Open Access Monitoring: guidelines and recommendations for research organisations and funders*. California: Science Europe, 2021. Disponível em: <https://www.scienceeurope.org/media/cqllmhz0/se-oamonitoring-briefing-paper-2021.pdf>. Acesso em: 2 jun. 2021.

SCIENTIFIC ELECTRONIC LIBRARY ONLINE. Além do SciELO, a FAPESP promove novas iniciativas em acesso aberto. *SciELO em Perspectiva*, [São Paulo], 15 jul. 2013. Disponível em: <https://blog.scielo.org/blog/2013/07/15/44/>. Acesso em: 12 jan. 2021.

SHINTAKU, M.; LANNÉ, S. C. da S. Dados abertos e a Fundação de Amparo à Pesquisa do Estado de São Paulo. In: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org.). *Tópicos sobre dados abertos para editores científicos*. São Paulo: ABEC, 2020. p. 87-96. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 15 mar. 2021.

SWAN, A. *The Open Access citation advantage: studies and results to date*. [S. l., 2010]. Disponível em: <https://eprints.soton.ac.uk/268516/>. Acesso em: 25 jun. 2020.

TONINI, A. M.; SANTANA, P. H. de A.; SCHERRER, T. M. A Ciência Aberta no CNPq: acesso aberto aos dados e resultados de pesquisas. In: SHINTAKU, M.; SALES, L. F.; COSTA, M. (org.). *Tópicos sobre dados abertos para editores científicos*. São Paulo: ABEC, 2020. p. 203-214. DOI 10.21452/978-85-93910-04-3. Disponível em: https://www.abecbrasil.org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf.

org.br/arquivos/Topicos_dados_abertos_editores_cientificos.pdf. Acesso em: 12 mar. 2021.

VAN LEEUWEN, T. N.; TATUM, C.; WOUTER, P. F. Exploring possibilities to use bibliometric data to monitor gold open access publishing at the national level. *Journal of the Association for Information Science and Technology*, New York, v. 69, n. 9, p. 1161-1173, 2018. DOI 10.1002/asi.24029. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/asi.24029>. Acesso em: 25 jun. 2020.

VAN DE SOMPEL, H.; LAGOZE, C. The Santa Fe Convention of the open archives initiative. *D-Lib Magazine*, [s. l.], v. 6, n. 2, 2000. DOI 10.1045/february2000-vandesompel-oai. Disponível em: <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>. Acesso em: 20 jul. 2021.

WEITZEL, S. da R. O mapeamento dos repositórios institucionais brasileiros: perfil e desafios. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 24, n. 54, p. 105-123, 2019. DOI 10.5007/1518-2924.2019v24n54p105. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2019v24n54p105/38092>. Acesso em: 9 ago. 2021.

WEITZEL, S. da R. As novas configurações do Acesso Aberto: desafios e propostas. *RECIIS – Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, Rio de Janeiro, v. 8, n. 2, p. 65-75, 2014. DOI 10.3395/reciis.v8i2.447. Disponível em: <https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/447/1102>. Acesso em: 12 maio 2021.

XAVIER, R. F. Evolução das plataformas de acesso aberto brasileiras: propriedades e perspectivas. *Ciência da Informação*, Brasília, DF, v. 48, n. 3, p. 497-500, set./dez. 2019. Disponível em: <https://revista.ibict.br/ciinf/article/view/4949/4469>. Acesso em: 16 dez. 2020.

XIA, J.; MYERS, R. L.; WILHOITE, S. K. Multiple open access availability and citation impact. *Journal of Information Science*, Amsterdam, v. 37, n. 1, p. 19-28, dez. 2010. DOI 10.1177/0165551510389358. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0165551510389358?journalCode=jisb>. Acesso em: 12 maio 2021.



CAPÍTULO 3

Explorando a relação entre *rankings* de universidades e o uso responsável de métricas e indicadores

Solange Maria dos Santos

Rankings de universidades: antecedentes, impactos, alcances e limites

Os *rankings* universitários são instrumentos comparativos destinados à avaliação pública de universidades baseadas em métricas. Esses *rankings* classificam e estabelecem uma ordem hierárquica que promove relações cada vez mais competitivas entre essas instituições.

Nas duas últimas décadas, os *rankings* de universidades consolidaram sua posição no cenário global da educação superior (Hannan; Ringel, 2023; Kaidesoja, 2022). Eles cresceram significativamente em número (Altbach, 2012; Santos, 2018), exercem níveis consideráveis de influência (Hazelkorn, 2011) e recebem atenção de uma variedade de audiências tão diversas como: jornalistas, formuladores de políticas, consultores, gestores universitários, estudantes e comunidade científica (Brankovic; Ringel; Werron, 2018).

No entanto, ao mesmo tempo, um grande número de estudiosos no campo do ensino superior e de outras áreas das ciências sociais identificaram não apenas sérios vieses e problemas metodológicos na elaboração dos *rankings* universitários, mas também documentaram vasta gama de consequências não intencionais dessas classificações que são contraproducentes para a pesquisa e a educação universitária (Brink, 2018; Erkkilä; Piironen, 2021; Fernandez-Cano *et al.*, 2018; Gadd, 2020; Hallonsten, 2021; Hannan;

Ringel, 2023; Hazelkorn, 2011; Kaidesoja, 2022; Marginson, 2014; Salmi; Saroyan, 2007).

Essa situação apresenta um paradoxo, pois, embora os *rankings* de universidade tenham aumentado sua influência e seu impacto ao longo do tempo, um número cada vez maior de pesquisadores reconhece que os problemas metodológicos e as consequências negativas não intencionais dessas classificações superaram seus benefícios potenciais. A despeito disso, muitos são contrários ao abandono total dos *rankings* universitários (Kaidesoja, 2022).

Antecedentes dos rankings de universidades

Os *rankings* universitários, que hoje são parte estabelecida do cenário global do ensino superior, não são um fenômeno novo. A história do surgimento e da evolução dessas classificações revela que a ideia de ranquear as instituições com base em alguma apreciação de sua qualidade surgiu nos Estados Unidos há mais de um século. Antecedentes dessas classificações remontam a 1870, quando uma comissão do US Bureau of Education começou a publicar relatórios estatísticos anuais que também incluíam uma classificação de instituições (Kehm, 2020; Salmi; Saroyan, 2007; Webster, 1986).

Listas informais de universidades americanas eram ocasionalmente publicadas, mas a ideia de classificar as instituições com base em alguma apreciação de sua qualidade ganhou força em 1910, com a proposta do psicólogo norte-americano e professor da Universidade da Pensilvânia James McKeen Cattell (Webster, 1986). Em 1925, Raymond Hughes, professor de química e posteriormente vice-reitor da Universidade de Miami, publicou um *ranking* baseado nas opiniões de um grupo de especialistas. A tendência se consolidou a partir de 1959 e, sobretudo, a partir da década de 1980, estendendo-se à graduação. Como bem caracteriza Kehm (2020), esses *rankings*, ainda que organizados por acadêmicos, não tiveram impactos além dos âmbitos nos quais foram elaborados.

De fato, o primeiro *ranking* nacional de universidades que a literatura registra¹ foi uma classificação comercial de um veículo de mídia, o US News & World foi publicado em 1983 pelo US News & World Report como um relatório de avaliação das melhores faculdades americanas (Dill, 2006). Foi com o US News & World Report Ranking que, pela primeira vez, a informação sobre a avaliação das universidades tornou-se facilmente acessível a alunos de graduação em potencial e seus pais, promovendo, assim, algumas mudanças nos hábitos dos consumidores de informação sobre avaliação da qualidade das universidades. Desde então, a ideia de produzir *rankings* foi seguida por outros veículos de comunicação com a publicação de classificações de âmbito nacional em outros países².

Ainda que a publicação do US News seja considerada um marco como primeira classificação formal, nesse período, a sociedade – em geral – e a comunidade científica – em especial – deram pouca atenção ao fato. Foi somente no começo do século XXI que os *rankings* de universidades começaram a proliferar e atrair considerável atenção – a partir da publicação do Academic Ranking of World Universities (ARWU) Universidade pela Shanghai Jiao Tong, em 2003. O ARWU foi criado no contexto do Projeto 985, iniciado em 1998 para o estabelecimento de universidades de padrão mundial – *world-class universities* – na China, com o objetivo de estabelecer um padrão de comparação (*benchmark*) da Universidade Shanghai Jiao Tong com as universidades estrangeiras (Santos, 2015; Santos, 2018).

O fato é que a chegada do Shanghai ARWU em 2003 deflagrou uma reação em cadeia com o surgimento de diversas classificações globais de universidades. O ano de 2004 viu o surgimento³ do World University Rankings publicado pelo *Times Higher Education Supplement* (THE) e pela

1 No Brasil há registro de um *ranking* de universidades publicado já em 1982. Trata-se do *ranking* Melhores Faculdades do Brasil (RMFB) publicado pela *Playboy Magazine* (Gonçalves, 2017).

2 Para mencionar alguns: os britânicos, *British Times Higher Education* e *The Guardian*; os alemães, *ZEIT* e *Der Spiegel*; o francês *Nouvel Observateur*; o irlandês *Sunday Times*; o italiano *La Repubblica*; o russo *Russian Finance*; o canadense *Maclean's*; o espanhol *El Mundo*; os mexicanos *Reforma* e *El Universal*; o chileno *El Mercurio*; o brasileiro *Folha de S. Paulo*; e provavelmente muitos outros ao redor do mundo (Kehn, 2020; Calderón; França; Gonçalves, 2017).

3 Em 2004 também foi lançado o Ranking Web of Universities, ou Webometrics Rankings, que a princípio media apenas a presença das Instituições de Ensino Superior (IES) na *web*.

empresa de consultoria de ensino superior Quacquarelli Symonds (QS). Em setembro de 2010, após o rompimento da parceria na elaboração do THE (2004 a 2009)⁴, a QS decidiu seguir utilizando a metodologia original e produzir, de maneira independente, seu próprio *ranking*, o QS University Rankings. U-Multirank, iniciado pela União Europeia e agora produzido por um consórcio de *think tanks* e institutos de pesquisa, e o CWTS Leiden Ranking do Centro de Estudos de Ciência e Tecnologia formam o segundo nível. A essas classificações, consideradas pioneiras ou de maior notoriedade, seguiram-se outros *rankings* globais, regionais, nacionais, por assunto e especializados. No geral, o ensino superior global é hoje densamente preenchido por *rankings* universitários, bem como por organizações como o Observatório International Ranking Expert Group (IREG), que se dedicam à promoção de *rankings* universitários (Santos, 2018).

Todos os anos, para atender às necessidades de informação dos distintos públicos interessados nos *rankings* de universidades, vários desses sistemas de classificação são desenvolvidos, atualizados e publicados por diversos tipos de instituições – revistas e jornais eminentes, institutos de pesquisas, sociedades profissionais, organizações governamentais, e consultorias independentes – e seus resultados geram uma pressão significativa, tanto sobre as universidades ao redor do mundo quanto sobre ministérios e instituições responsáveis por elaborar políticas nacionais e por acompanhar o desempenho das universidades nos *rankings* globais (Usher; Savino, 2007). Segundo Thiengo (2018), essa pressão exercida pelos *rankings* se deve muito à capacidade que essas classificações têm de expor, culpar e até envergonhar os países e as instituições que não alcançam boas posições.

Um outro desdobramento importante do êxito dos *rankings* foi a criação de um “mercado de *rankings*”. A produção de um *ranking* tornou-se uma grande oportunidade de negócio para muitas empresas⁵, principalmente,

4 Este *ranking* é divulgado em parceria com a US News & World Report de Washington, D.C., o *Chosun Ilbo*, jornal da Coreia do Sul, o *The Sunday Times* do Reino Unido e o francês *Nouvel Observateur*.

5 Foram criados novos empreendimentos e serviços especializados voltados para a produção de *rankings*, como aplicativos para *smartphones*, permitindo comparar universidades e *rankings*; realização de conferências, palestras e seminários; serviços de consultoria de educação superior, tanto para orientar sobre escolha de universidades, quanto para orientar institui-

para as organizações de mídia, responsáveis pela produção de *rankings* em diversos países. Desse modo, além da competição entre as universidades, passou a ocorrer também uma acirrada competição no próprio campo da elaboração desses instrumentos, que tem motivado o surgimento e a proliferação de uma diversidade de *rankings* orientados à classificação, à avaliação e à ordenação de Instituições de Ensino Superior (IES). Nesse cenário, junto a *rankings* clássicos e de maior prestígio – como o ARWU, o THE e o QS –, há também *rankings* ainda pouco conhecidos centrados em avaliar o desempenho em inovação, empregabilidade, sustentabilidade ambiental, dentre outras perspectivas e dimensões da atuação das universidades (Santos, 2018).

Além dessa diversidade de *rankings* globais, há também uma variedade de sistemas de classificação em nível nacional desenvolvidos por instituições locais. Estima-se que existam mais de 60 *rankings* nacionais sendo publicados regularmente em mais de 30 países (Inorms Research Evaluation Group, 2001; Santos, 2018).

Dessa forma, o monitoramento dessas iniciativas, assim como sua evolução ao longo do tempo, permite que instituições diversas possam avaliar a mudança do contexto que as envolve e instruir seu posicionamento.

Principais críticas aos rankings de universidades

É verdade que as tabelas de classificações e *rankings* universitários tentam capturar em um único índice as múltiplas atividades das universidades, tarefa bastante complexa, se não impossível. Diante disso, apesar da popularidade alcançada, os *rankings* têm sido alvo de intensas críticas ao longo das últimas duas décadas, não obstante, essas críticas não têm impedido a conquista de legitimidade e credibilidade desses convertidos instrumentos de avaliação, que seguem sendo vistos como referências de uma determinada excelência e qualidade (Hannan; Ringel, 2023; Kaidesoja, 2022).

ções sobre melhor posicionamento nos *rankings*; serviços de levantamento e produção de dados estatísticos e indicadores bibliométricos, dentre outros (Ball, 2014; Robertson; Olds, 2012; Thiengo, 2018).

Os *rankings* universitários são um tema frequente em debates científicos, em oficinas, em conferências, em artigos de revistas, volumes editados, monografias ou relatórios. No entanto, os críticos acadêmicos também usam domínios que permitem que eles abordem públicos mais diversos – e maiores –, como eventos presenciais – onde podem se envolver com os classificadores e outras partes interessadas diretamente –, a mídia de massa – na qual publicam artigos de opinião ou comentários sobre *rankings* em entrevistas –, *sites* pessoais, *blogs* e mídias sociais (Hannan; Ringel, 2023).

Os principais críticos chamam a atenção para os múltiplos efeitos negativos e deficiências metodológicas dos *rankings* universitários, que foram sintetizados por Hannan e Ringel (2023) como segue:

- **Efeitos negativos.** A propensão dos *rankings* universitários para produzir novas desigualdades e consolidar as existentes é, talvez, o efeito mais criticado. Embora os *rankings* universitários criem uma aura de neutralidade, eles são, na verdade, “[...] tecnologias político-ideológicas de valoração e hierarquização que operam de acordo com uma [...] lógica de inclusão e exclusão” (Amsler; Bolsmann, 2012, p. 284). As desigualdades são identificadas tanto em termos materiais quanto simbólicos, e dizem respeito ao indivíduo e ao nível institucional (Hamann, 2016). Estudos que se concentram nas desigualdades materiais no nível individual sugerem, por exemplo, que os *rankings* aumentam as mensalidades e, assim, impedem que estudantes de baixa renda se candidatem a universidades de “elite” (Chu, 2021). No nível institucional, diz-se que os *rankings* levam à monopolização do financiamento da pesquisa, promovendo um sistema no qual poucas universidades recebem a maioria dos recursos (Münch, 2014). Outros criticam que os *rankings* (re)produzem desigualdades simbólicas ao criarem “geografias multiescalares de reputação institucional” (Collins; Park, 2016, p. 116), que dizem incutir uma “busca sem fim pelo status de classe mundial” (Hazelkorn, 2011, p. 24) e “legitimam a colocação de uma instituição na hierarquia global” (Kauppinen *et al.*, 2016);
- **Oportunismo.** Outro tipo de crítica chama a atenção para o aumento do oportunismo provocado pelos *rankings* universitários. A principal preocupação é que, por serem classificadas, as universidades tenham

desenvolvido uma obsessão com a reputação e adaptem seu comportamento para se adequar aos critérios de avaliação com os quais são comparadas (Espeland; Sauder, 2007). Cada vez mais, as universidades são vistas motivando a equipe acadêmica ao ceder a um “hábito proliferante de conectar recompensas financeiras a posições de classificação [...]” (Hallonsten, 2021, p. 14). Esses tipos de atividades, argumentam os críticos, desviam as universidades de sua missão principal de fornecer apoio institucional para a pesquisa básica e o ensino (Collins; Park, 2016; Van Houtum; Van Uden, 2022). Nesse sentido, um artigo publicado no *site* University World News afirma: “na ausência de padrões globais ou regionais ou harmonização da coleta e definição de dados, as várias maneiras pelas quais as instituições jogam o jogo dos *rankings* persistirão” (Calderon, 2020);

- **Influência externa.** Um terceiro efeito negativo dos *rankings* universitários se deve à sua influência externa. A seguinte citação resume a premissa básica desta crítica: “as classificações acabam por colonizar a educação e a ciência [...] e estas impõem sua própria lógica de produção de diferenças hierárquicas à prática da pesquisa e do ensino” (Münch, 2013, p. 201). Nessa perspectiva, os *rankings* promovem regimes intrusivos de responsabilização na academia, incluindo “novas configurações de conhecimento e poder” (Shore; Wright, 2015, p. 430), que equivalem a violações da autonomia profissional da ciência e da primazia do julgamento profissional (Hamann; Schmidt-Wellenburg, 2020; Mollis; Marginson, 2002). Em outras palavras, a natureza quantitativa dos *rankings* pode interferir na capacidade das universidades de cumprir sua missão acadêmica. A preocupação é que as universidades estejam sendo forçadas a se concentrar em métricas específicas para obter uma classificação mais alta, em detrimento de outras atividades importantes, como a pesquisa e o ensino (Hannan; Ringel, 2023).

Deficiências metodológicas dos rankings universitários

- **Dados fragmentados.** Um argumento comum contra os *rankings* é que eles dividem e transformam conceitos como “qualidade de ensino” ou “desempenho em pesquisa” em medidas que são muito simplistas e

falham em fazer jus à complexa realidade que pretendem retratar. Em sua essência, esse tipo de crítica questiona se uma única medida pode realmente refletir o desempenho geral em uma variedade de atividades de produção relacionadas ao ensino e à pesquisa (Johnes, 2018). Os *rankings* são criticados por agregar uma variedade de indicadores em pontuações gerais e por reivindicar um nível de precisão holística que não só é irrealista do ponto de vista científico, mas também incentiva o uso deles “[...] como um proxy preguiçoso para qualidade, independentemente das falhas” (Gadd, 2020, p. 523). Sendo assim, de acordo com Gapper (2021), “os *rankings* de universidades são inerentemente questionáveis porque eles usam vários pedaços de dados [...] para chegar a um número geral”;

- **Falta de transparência.** Os críticos também apontam para a falta de transparência em relação aos indicadores e aos dados usados pelos *rankings*. Essas deficiências comprometem a credibilidade científica e a integridade destes (Marginson, 2014; Surappa, 2016). Por exemplo, um estudo sugere que o *ranking* do US News “[...] não é totalmente claro quando se trata dos dados e, na verdade, suprime os valores de alguns dos atributos” (Bougnol; Dulá, 2015, p. 864). Em um estudo que compara cinco *rankings*, os autores destacam que eles “[...] falham em fornecer uma justificativa teórica ou empírica para as medidas selecionadas e os pesos utilizados para calcular suas classificações” (Dill; Soo, 2005, p. 506). Isso posto, segundo Holmes (2020), “os *rankings* da THE são singularmente opacos: eles combinam onze indicadores em três clusters, então é impossível para o leitor descobrir exatamente por que uma universidade está indo tão bem ou tão mal em ensino ou em pesquisa”;
- **Validade e confiabilidade dos métodos e dados.** A lógica básica é que os *rankings* universitários não entregam o nível de precisão que prometem; suas medições não são apenas vistas como simplistas, mas como fundamentalmente falhas. Os *rankings* universitários são considerados medidas “[...] não suficientemente confiáveis de qualidade de desempenho e não fornecem a base para atividades verificáveis de melhoria” (Leiber, 2017, p. 47). Portanto, não é surpreendente que

eles sejam acusados de fornecer resultados não confiáveis, até mesmo arbitrários. “Os resultados dos distintos rankings diferem significativamente, porque as metodologias subjacentes são diferentes” (Schmoch, 2014, p. 152). O fato de alguns *rankings* universitários se basearem em pesquisas de opinião levanta preocupações específicas sobre sua confiabilidade e validade (Dill; Soo, 2005). Nesses casos, os críticos sugerem o uso de outros indicadores mais adequados (Bougnol; Dulá, 2015; Kroth; Daniel, 2008).

Embora continuamente alvo de debates que enfatizam efeitos negativos ou deficiências metodológicas, os *rankings* universitários parecem ser bastante resilientes a essas críticas (Hannan; Ringel, 2023). Nesse sentido, os responsáveis pela produção e publicação dos principais *rankings* alegam que os resultados destes, conforme publicados, são apenas uma interpretação dos dados e alertam que, devido às suas limitações, não se deve usar posições e tabelas de classificação como padrão ou métrica isolada (Baty, 2010). Segundo o autor, os *rankings* são melhores quando usados como ferramentas de tomada de decisão em conjunto com outros indicadores e dados (Baty, 2010).

Os *rankings* de universidades, assim como outros dispositivos apoiados em métricas, têm sido utilizados há muito tempo como um indicador de sucesso acadêmico e como uma forma de tomar decisões importantes. Contudo, à medida que a aferição do impacto se torna cada vez mais importante dentro da academia, tem surgido uma reação contra a dependência excessiva de métodos de avaliação puramente quantitativos.

Nesse cenário, o movimento pelo uso responsável de métricas e indicadores visa garantir que os números sejam usados juntamente com outras medidas para obter uma avaliação responsável e justa do real impacto da pesquisa. A expectativa é que esse movimento possa alcançar e impactar também a produção e o uso de *rankings* e tabelas de classificação de universidades.

Movimento pelo uso responsável de métricas e indicadores na avaliação da pesquisa

À medida que a pesquisa se institucionalizava nas universidades, a avaliação se tornou um instrumento cada vez mais complexo, projetado para selecionar pesquisadores, avaliar instituições ou conceder subsídios para projetos de pesquisa (Gingras, 2016). Nesse contexto, muitos esforços de avaliação concentram-se principalmente em mensurar o que pode ser facilmente identificável, como o número e o montante de concessões de financiamento, além do número e das citações de artigos publicados. No entanto, Moher e demais autores (2020) destacam que mesmo para aspectos prontamente mensuráveis, os critérios usados para avaliação e decisões podem variar entre ambientes e instituições e não são necessariamente aplicados de forma consistente, mesmo dentro da mesma instituição.

Essa crescente insatisfação com os sistemas de avaliação na comunidade científica reflete uma percepção de que o atual sistema de incentivos e recompensas dos pesquisadores não está alinhado com as necessidades da sociedade (Moher *et al.*, 2020). Os regimes de avaliação acadêmica estabelecidos para quantificar e avaliar a qualidade da pesquisa, dos acadêmicos e das instituições têm sido amplamente criticados devido aos efeitos prejudiciais que exercem sobre os ambientes acadêmicos e sobre a própria produção de conhecimento.

A insatisfação da comunidade científica com os sistemas de avaliação e a necessidade de modificar os critérios empregados acabaram culminando numa variedade de pronunciamentos públicos muito críticos em relação ao uso indiscriminado de indicadores como o Fator de Impacto nos processos avaliativos. Esses posicionamentos sugerem, entre outras medidas, incorporar pontos de vista qualitativos, avaliação no nível do artigo – não centrada no Fator de Impacto do periódico – e também ampliação do espectro de indicadores a serem incorporados nas avaliações, mas também a recomendação do uso de métricas responsáveis robustas, transparentes, diversas e reflexivas.

A seguir, teremos a oportunidade de apresentar mais detalhadamente alguns dos principais posicionamentos da comunidade científica em favor do uso responsável de métricas e indicadores em processo avaliativos.

Posicionamento em prol do uso responsável de indicadores na avaliação da pesquisa

Como vimos, tem havido nos últimos anos um interesse crescente da comunidade científica em discutir e recomendar o uso responsável de indicadores na avaliação. Como parte dos esforços desse movimento, uma série de documentos com posicionamento de alto nível foram publicados por grandes grupos – por exemplo, o Manifesto de Leiden –, também por grupos menores ou individuais – por exemplo, Ioannidis e Khoury (2014); ou por iniciativas de periódicos (Time [...], 2016).

O advento de manifestos e declarações como a Declaração de San Francisco para Avaliação da Pesquisa (DORA), Manifesto de Leiden, o relatório *The Metric Tide* ou, mais recentemente, os Princípios de Hong Kong expuseram as várias distorções que o sistema de publicação e avaliação haviam gerado na vocação universalista da ciência e destacaram a necessidade de buscar as melhores práticas e reduzir os abusos no emprego de métricas e indicadores. A continuação é apresentada, em ordem cronológica: as principais declarações e posicionamentos da comunidade científica acerca de uso responsável de métricas e indicadores no que se refere à avaliação da pesquisa acadêmica.

Declaração sobre Avaliação da Pesquisa (DORA)

A DORA foi desenvolvida por um grupo de editores e representantes de editoras de revistas acadêmicas em 16 de dezembro 2012, durante a reunião anual da American Society for Cell Biology (ASCB) em San Francisco e, desde então, se tornou uma iniciativa mundial que abrange todas as disciplinas. O grupo elaborou um conjunto de recomendações e a declaração inclui princípios fundamentais para financiadores, editores, sociedades profissionais, instituições e pesquisadores.

Embora contenha várias recomendações, o principal alvo da DORA é o uso abusivo do Fator de Impacto. O Fator de Impacto⁶ é frequentemente usado como o parâmetro na comparação da produção científica de indivíduos e instituições. Essa métrica foi originalmente concebida como uma ferramenta para auxiliar bibliotecários no processo de aquisição de periódicos mais importantes para melhor atender às suas comunidades dentro de uma disciplina específica, mas que não tem qualquer utilidade como medida da qualidade científica das pesquisas publicadas. Logo, é conceitualmente errado utilizar o Fator de Impacto para avaliar a importância de um artigo para fins de contratação, promoção, avaliação de pesquisadores individuais ou grupos de pesquisa.

Nesse sentido, a DORA defende claramente que qualquer outro uso do Fator de Impacto é mau uso.

6 As deficiências do Fator de Impacto enquanto ferramenta para a avaliação da pesquisa são amplamente conhecidas e muito bem documentadas na literatura: a) algumas distribuições de citações dentro de cada periódico podem ser muito desiguais; b) as propriedades do Fator de Impacto são dependentes do campo: trata-se de um conjunto de muitos tipos de artigos, muito diferentes, incluindo artigos, resultado, pesquisa primária, resenhas e editoriais; c) Fatores de Impacto podem ser distorcidos por políticas e práticas editoriais questionáveis; e d) os dados usados para calcular o Fator de Impacto não são transparentes nem abertamente acessíveis ao público, dentre outros tipos de disfunção (Declaration on Research Assessment, 2012).

Quadro 1 - Recomendações da DORA a agências de fomento

RECOMENDAÇÕES DA DORA	
Geral	1. Não usar métricas baseadas em revistas, como Fatores de Impacto, como medida substitutiva da qualidade de artigos de pesquisa individuais para avaliar as contribuições de um cientista individual, ou em decisões relativas a contratações, promoções ou financiamentos.
Para agências de fomento	2. Explicitar os critérios usados na avaliação da produtividade de solicitantes de bolsas e enfatizar claramente, em especial para pesquisadores em início de carreira, que o conteúdo científico de um artigo é muito mais importante que a métrica ou a identidade da revista em que é publicado. 3. Para fins de avaliação da pesquisa, considerar o valor e o impacto de todos os resultados – inclusive conjuntos de dados e <i>software</i> – além das publicações, e considerar um amplo espectro de medidas de impacto, inclusive indicadores qualitativos de impacto da pesquisa, tais como as influências em políticas e práticas. 4. Explicitar os critérios usados nas decisões de contratação, concessão de estabilidade (<i>tenure</i>), e promoção, enfatizando claramente, em especial para pesquisadores novatos, que o conteúdo científico de um artigo é muito mais importante que a métrica ou a identidade da revista em que é publicado.
Para editoras	5. Reduzir drasticamente a ênfase no Fator de Impacto como recurso promocional, idealmente deixando promovê-lo, ou apresentando-o no contexto de uma diversidade de métricas referentes a revistas – p. ex.: Fator de Impacto quinquenal, EigenFactor, SCImago, índice-h, prazos editoriais e de publicação etc. – que fornecem uma visão mais rica do desempenho da revista. 6. Disponibilizar uma variedade de métricas referentes a artigos para encorajar a mudança em direção a avaliações baseadas no conteúdo científico de cada artigo, em vez de na métrica da revista em que é publicado. 7. Encorajar práticas responsáveis de autoria, ademais do fornecimento de informações sobre as contribuições específicas de cada autor. 8. Quer uma revista seja de acesso aberto, quer baseada em assinaturas, remover todas as limitações sobre o reuso em listas de referências em artigos de pesquisa e disponibilizá-las de acordo com a Creative Commons Public Domain Dedication. 9. Remover ou reduzir as restrições ao número de referências em artigos de pesquisa e, quando for adequado, tornar obrigatória a citação da literatura primária no lugar de resenhas para dar crédito ao grupo, ou grupos, que primeiro relatou uma descoberta.
Para organizações que fornecem métricas	10. Ser aberto e transparente deixando claros os dados e métodos usados para calcular todas as métricas. 11. Fornecer os dados sob uma licença que permita reuso irrestrito, e acesso computacional aos dados, quando possível. 12. Deixar claro que a manipulação inadequada de métricas não será tolerada; explicitar os critérios do que constitui manipulação inadequada e quais medidas serão tomadas para combatê-la. 13. Levar em conta a diversidade nos tipos de artigo – p. ex.: revisões em comparação com artigos de pesquisa –, e nas áreas do conhecimento quando métricas são usadas, agregadas ou comparadas.

RECOMENDAÇÕES DA DORA	
Para os pesquisadores	14. Ao participar de comissões deliberativas sobre o financiamento, a contratação, a concessão de estabilidade (<i>tenure</i>) ou a promoção, fazer avaliações baseadas no conteúdo científico, em vez de em métricas de publicação. 15. Quando apropriado, citar a literatura primária em que observações são relatadas, em vez de resenhas, a fim de dar crédito a quem o merece. 16. Usar um espectro de métricas de artigo e indicadores em currículos, como evidência do impacto de artigos individuais publicados e outros resultados da pesquisa. 17. Questionar práticas de avaliação da pesquisa baseadas impropriamente em Fatores de Impacto, com o fito de promover e ensinar a prática mais adequada, centrada no valor e na influência de resultados de pesquisa específicos.

Fonte: adaptado de Declaration on Research Assessment (2012).

A DORA também oferece a possibilidade de “assiná-la” – transformando-a em uma possível declaração pública que confirma que um indivíduo ou instituição está comprometida com os princípios nela contidos. Em julho de 2020, a DORA contava com mais de 1,9 mil intuições signatárias e mais de 16 mil assinaturas individuais.

Após a DORA, outros manifestos e posicionamentos internacionais publicados seguiram direção semelhante.

Manifesto de Leiden

O Manifesto de Leiden para Métricas de Pesquisa foi publicado pela primeira vez na forma de comentário no periódico *Nature*, em 2015. Cinco especialistas liderados por Diana Hicks, professora da Escola de Políticas Públicas do Georgia Institute of Technology, e Paul Wouters, diretor do CWTS da Universidade de Leiden, que reuniram o pensamento coletivo dos participantes da conferência da qual a iniciativa se originou⁷.

Os dez princípios propostos no manifesto não são novidade para os pesquisadores das áreas de bibliometria e cientometria, mas esses especialistas normalmente não estão presentes quando os avaliadores se reportam aos gestores universitários, que também não são especialistas no tema. Além

7 Science and Technology Indicators Conference (STI) realizada entre 3 e 5 de setembro de 2014 em Leiden, na Holanda, pelo CWTS em colaboração com a European Network of Indicator Developers (ENID). Disponível em: <http://sti2014.cwts.nl/Home>.

disso, a literatura necessária para contestar ou questionar o uso inadequado de indicadores nas avaliações encontra-se, normalmente, publicada em periódicos especializados e nem sempre de modo acessível.

Desse modo, o Manifesto de Leiden sintetiza algumas das melhores práticas de avaliação da pesquisa baseada em métricas, para que os pesquisadores avaliados possam confiar em seus avaliadores e para que os avaliadores possam confiar em bem empregar os indicadores (Hicks *et al.*, 2015).

Quadro 2 - Dez princípios propostos pelo Manifesto de Leiden

PRINCÍPIOS DO MANIFESTO DE LEIDEN	
1	A avaliação quantitativa deve apoiar especialistas qualitativos na avaliação. Os indicadores quantitativos podem corrigir tendências enviesadas da avaliação por pares e facilitar a deliberação. Nesse sentido, devem fortalecer a revisão por pares já que emitir julgamentos sobre colegas é difícil sem apoio de uma série de informações relevantes. No entanto, os avaliadores não devem ceder à tentação de basear suas decisões apenas em números. Os indicadores não devem substituir o juízo informado. Os tomadores de decisão têm plena responsabilidade por suas avaliações.
2	Medir o desempenho em relação às missões de pesquisa da instituição, do grupo ou do pesquisador. Os objetivos de um programa de pesquisa devem ser indicados no início, e os indicadores utilizados para avaliar o desempenho devem estar claramente vinculados aos objetivos. A escolha dos indicadores e de como eles são utilizados deve levar em conta os contextos socioeconômicos e culturais de maneira mais ampla. Os cientistas têm diversas missões de pesquisa. A pesquisa que avança as fronteiras do conhecimento acadêmico difere da pesquisa que é focada em proporcionar soluções para os problemas da sociedade. A avaliação pode ser baseada em méritos relevantes para as políticas públicas, para a indústria ou para os cidadãos em geral, em vez de méritos baseados em noções acadêmicas de excelência. Não existe um modelo único de avaliação que se aplique a todos os contextos.
3	Proteger a excelência em pesquisas relevantes localmente. Em muitas partes do mundo, a excelência da pesquisa é associada à publicação no idioma inglês. O Fator de Impacto é calculado na Web of Science (WoS), que indexa principalmente os periódicos com base nos Estados Unidos e em língua inglesa. Este viés é particularmente problemático para as ciências sociais e humanidades, áreas mais orientadas para a pesquisa de temas regionais e nacionais. Muitas outras áreas possuem uma dimensão nacional ou regional. O pluralismo e a relevância para a sociedade tendem a ser anulados quando se produzem artigos de interesse para os defensores do alto impacto: os periódicos em inglês. Os indicadores baseados nos periódicos de qualidade publicados em outros idiomas que não o inglês devem identificar e privilegiar as áreas de pesquisa de interesse local.

PRINCÍPIOS DO MANIFESTO DE LEIDEN	
4	Manter a coleta de dados e os processos analíticos abertos, transparente e simples. A construção das bases de dados necessárias para a avaliação deve observar regras claramente definidas e fixadas antes da conclusão da pesquisa. Esta era a prática comum entre os grupos acadêmicos e comerciais que desenvolveram metodologias de avaliação bibliométrica ao longo de muitas décadas. Tais grupos referenciaram protocolos publicados na literatura revisada por pares. Esta transparência possibilitou o escrutínio das metodologias. Por exemplo, em 2010, o debate público sobre as propriedades técnicas de um importante indicador utilizado por um dos nossos grupos – o CWTS, da Universidade de Leiden, na Holanda – levou a uma revisão no cálculo desse indicador (6). Os novos operadores do setor privado devem seguir os mesmos padrões; ninguém deve aceitar avaliações saídas de uma caixa-preta.
5	Permitir que os avaliados verifiquem os dados e as análises. Para garantir a qualidade dos dados, todos os pesquisadores incluídos em estudos bibliométricos deveriam poder verificar se suas produções foram corretamente identificadas. Todos os que dirigem e administram os processos de avaliação devem assegurar a precisão dos dados, por meio de verificação própria ou auditoria de terceiros. As universidades poderiam implementar esse princípio em seus sistemas de informação sobre a pesquisa, o que deveria ser um princípio norteador na seleção de fornecedores desses sistemas. A coleta e o processamento de dados precisos e de alta qualidade demandam tempo e dinheiro e devem ser considerados no orçamento institucional.
6	Considerar as diferenças entre áreas nas práticas de publicação e citação. A melhor prática de avaliação é selecionar um conjunto de possíveis indicadores e permitir que as distintas áreas escolham aqueles que lhes são mais adequados. Há alguns anos, um grupo europeu de historiadores recebeu uma classificação relativamente baixa em uma avaliação nacional por pares, porque escreviam livros em vez de artigos em revistas indexadas na WoS. Esses historiadores tiveram o azar de fazer parte de um departamento de Psicologia. Historiadores e cientistas sociais precisam que os livros e a literatura publicada no idioma nacional sejam incluídos na contagem de publicações; já os cientistas da computação esperam que seus trabalhos apresentados em eventos e conferências sejam levados em conta. Os valores de citações variam por área: os periódicos mais bem avaliados em Matemática têm Fator de Impacto por volta de 3; já as revistas mais bem avaliadas em Biologia Celular tem Fator de Impacto em torno de 30. Portanto, é necessário o uso de indicadores normalizados, e o método de normalização mais confiável é baseado em percentuais: cada artigo é ponderado segundo o percentual a que pertence na distribuição de citações em sua área – os melhores 1%, 10% ou 20%, por exemplo. Uma única publicação altamente citada melhora ligeiramente a posição de uma universidade em um <i>ranking</i> baseado em indicadores percentuais, mas pode impulsionar a universidade de uma posição mediana para as primeiras posições em um <i>ranking</i> baseado em médias de citação.
7	Basear a avaliação de pesquisadores individuais no juízo qualitativo da sua carreira. Quanto mais idade você tem, maior será o seu Índice h, mesmo que não publique novos artigos. O Índice h varia por área: os pesquisadores das ciências da vida chegam ao topo com 200; os físicos com 100 e os cientistas sociais com 20 a 30. Depende da base de dados: há pesquisadores em Ciência da Computação que têm um Índice h de cerca de 10 na WoS, mas de 20 a 30 no Google Scholar. Ler e julgar o trabalho de um pesquisador é muito mais adequado do que depender de um número. Mesmo quando se compara grande número de pesquisadores, uma abordagem que considere informações diversas sobre o conhecimento, a experiência, as atividades e a influência de cada indivíduo é a melhor.

PRINCÍPIOS DO MANIFESTO DE LEIDEN	
8	Evitar objetividade mal colocada e falsa precisão. Indicadores de ciência e tecnologia são propensos à ambiguidade conceitual e à incerteza, e demandam fortes suposições que não são universalmente aceitas. O significado das contagens de citações, por exemplo, tem sido amplamente discutido. Assim, a melhor prática de avaliação utiliza indicadores múltiplos para fornecer uma imagem mais robusta e plural da pesquisa. Se as incertezas e os erros podem ser quantificados, esta informação deve acompanhar os valores dos indicadores publicados, usando barras de erro, por exemplo. Se isso não for possível, os produtores de indicadores deveriam, pelo menos, evitar a falsa precisão. Por exemplo, o Fator de Impacto de revistas é publicado com três casas decimais para evitar empates. No entanto, dada a ambiguidade conceitual e a variabilidade aleatória das contagens de citações, não faz sentido distinguir as revistas com base em diferenças mínimas do Fator de Impacto. Evite a falsa precisão: apenas uma casa decimal se justifica.
9	Reconhecer os efeitos sistêmicos da avaliação e dos indicadores. Os indicadores mudam o sistema da pesquisa por meio dos incentivos que estabelecem. Estes efeitos devem ser previstos. Isto significa que um conjunto de indicadores é sempre preferível – um único indicador convida a burlas ou a desvios do objetivo, em que a medida se torna um fim em si. Por exemplo, na década de 1990, a Austrália financiou a pesquisa universitária por intermédio de uma fórmula baseada, sobretudo, no número de artigos publicados pelas instituições. As universidades poderiam calcular o “valor” de um artigo em uma revista revisada por pares; em 2000, o valor era de estimados AUS\$ 800 – em torno de US\$ 480 – para o financiamento da pesquisa. Previsivelmente, o número de artigos publicados por pesquisadores australianos subiu, mas em revistas menos citadas, sugerindo uma queda na qualidade dos artigos.
10	Examinar e atualizar os indicadores regularmente. A missão da pesquisa e os objetivos da avaliação mudam, e o próprio sistema de pesquisa evolui junto. Medidas que anteriormente eram úteis se tornam inadequadas e surgem novos indicadores. Os sistemas de indicadores têm de ser revistos e talvez modificados. Percebendo os efeitos de sua fórmula simplista, em 2010, a Austrália introduziu a iniciativa “Excelência na Pesquisa para a Austrália” (Excellence in Research for Australia), mais complexa e com ênfase na qualidade.

Fonte: adaptado de Hicks *et al.* (2015).

Os pontos 6 a 10 do Manifesto de Leiden fornecem exemplos relevantes sobre a variação das métricas chamando atenção para o fato de que métricas flutuam e isso é normal. Essas características raramente são consideradas quando se avalia o desempenho de pesquisadores ou quando se analisa a classificação de universidades nos *rankings* internacionais.

The metric tide

O relatório *The metric tide: report of the independent review of the role of metrics in research assessment and management*, publicado em julho de 2015, foi coordenado pelo professor James Wilsdon, professor de Ciência e Democracia da Universidade de Sussex, apoiado por um grupo independente

e multidisciplinar de especialistas em cientometria, financiamento de pesquisa, política de pesquisa, publicação, gestão universitária e administração de pesquisa.

O relatório explora o uso de métricas em diferentes disciplinas e avalia seu potencial de contribuição para o desenvolvimento da excelência, da avaliação e do impacto da pesquisa. Além disso, analisa a aplicabilidade e os efeitos negativos ou não intencionais das métricas em vários aspectos da cultura de pesquisa, incluindo gerenciamento por desempenho, igualdade, diversidade e interdisciplinaridade.

O Grupo de Trabalho (GT) responsável pela elaboração do *The metric tide* identificou 20 recomendações que demandam atenção e ação das partes interessadas em todo o sistema de pesquisa do Reino Unido, mas que podem ser de muita utilidade para outros sistemas nacionais de pesquisa. O GT propõe ações nas seguintes áreas: apoiar a liderança eficaz, a governança e a gestão das culturas de pesquisa; melhorar a infraestrutura de dados que apoia o gerenciamento de informações de pesquisa; aumentar a utilidade dos dados e fontes de informação existentes; e promover o uso adequado de métricas na avaliação do Research Excellence Framework (REF)⁸.

8 O REF é o sistema do Reino Unido para avaliar a qualidade da pesquisa nas IES. O REF é realizado pelos quatro organismos de financiamento do ensino superior do Reino Unido: Research England, Scottish Funding Council (SFC), o Higher Education Funding Council for Wales (HEFCW), e o Department for the Economy (DfE). O REF atribui uma pontuação às instituições e o desempenho é utilizado para determinar a alocação de financiamento de pesquisa entre as IES.

Quadro 3 – Recomendações do relatório independente *The metric tide*

THE METRIC TIDE – RECOMENDAÇÕES		
OBJETIVOS	RECOMENDAÇÕES	DESTINATÁRIOS
Apoiar a liderança, a governança e o gerenciamento eficazes das culturas de pesquisa	1. A comunidade de pesquisa deve desenvolver uma abordagem mais sofisticada e diferenciada da contribuição e das limitações dos indicadores quantitativos. É necessário cuidado com a linguagem e a terminologia adotadas. O termo “métricas” não é mais adequado; o termo preferencial é “indicadores”, que reflete o reconhecimento de que os dados podem não ter relevância específica, mesmo que sejam úteis em um contexto geral.	IES, financiadores, gestores, pesquisadores
	2. No nível institucional, os líderes das IES devem desenvolver uma declaração clara de princípios sobre sua abordagem à gestão e à avaliação da pesquisa, incluindo o papel dos indicadores quantitativos. Com base nesses princípios, eles devem selecionar cuidadosamente indicadores quantitativos adequados aos seus objetivos e contextos institucionais. Nos casos em que as instituições usam tabelas de classificação, elas devem explicar por que as estão usando como um meio para alcançar fins específicos. Sempre que possível, indicadores alternativos que apoiem a igualdade e a diversidade devem ser identificados e incluídos. A comunicação clara da justificativa para a seleção de indicadores específicos e como eles serão usados como uma ferramenta de gerenciamento é fundamental. Como parte desse processo, as IES devem considerar assinar a DORA ou basear-se em seus princípios e adaptá-los aos seus contextos institucionais.	Dirigentes de instituições, dirigentes de pesquisa, gestores de IES
Apoiar a liderança, a governança e o gerenciamento eficazes das culturas de pesquisa	3. Os gestores e administradores de pesquisa devem defender esses princípios e o uso de métricas responsáveis em suas instituições. Eles devem prestar a devida atenção às implicações de igualdade e diversidade das opções de avaliação da pesquisa; envolver-se com especialistas externos; ajudar a facilitar uma infraestrutura de dados mais aberta e transparente; defender o uso de identificadores exclusivos, como ORCID IDs; trabalhar com financiadores e editores na interoperabilidade de dados; explorar indicadores para aspectos da pesquisa que eles desejam avaliar, em vez de usar os indicadores existentes apenas por estarem prontamente disponíveis; aconselhar os líderes seniores em métricas significativas para o seu contexto institucional ou departamental; e compartilhar boas práticas por órgãos setoriais.	Gestores, administradores de pesquisa, associações de gestores e administradores de pesquisa
	4. Os gerentes de RH e os painéis de recrutamento ou promoção nas IES devem ser explícitos sobre os critérios usados nas decisões de nomeação e promoção acadêmica. Esses critérios devem ser fundamentados em especialistas em julgamento e podem refletir tanto a qualidade acadêmica dos resultados quanto as contribuições mais amplas para políticas, indústria ou sociedade. Às vezes, os julgamentos podem ser útilmente guiados por métricas, se relevantes para os critérios em questão, e utilizados de forma responsável.	Gerentes de Recursos Humanos (RH), painéis de recrutamento e promoção

THE METRIC TIDE – RECOMENDAÇÕES		
OBJETIVOS	RECOMENDAÇÕES	DESTINATÁRIOS
Apoiar a liderança, a governança e o gerenciamento eficazes das culturas de pesquisa	5. Pesquisadores individuais devem estar atentos às limitações de indicadores específicos na maneira como apresentam seus próprios currículos e avaliam o trabalho dos colegas. Quando os indicadores padrão são inadequados, os pesquisadores individuais devem procurar uma variedade de fontes de dados para documentar e apoiar reivindicações sobre o impacto do seu trabalho.	Todos os pesquisadores
	6. Organizações como IES e financiadores de pesquisa devem desenvolver seus próprios princípios, em seus contextos específicos, sobre o uso de indicadores quantitativos na avaliação e na gestão da pesquisa e devem garantir que sejam claramente comunicados, fáceis de localizar e entender. Eles devem buscar abordagens para a coleta de dados que sejam transparentes, acessíveis e permitam maior interoperabilidade em uma diversidade de plataformas.	Órgãos de financiamento ao ensino superior, conselhos de pesquisa e outros financiadores de pesquisa
	7. Os provedores de dados, analistas e produtores de <i>rankings</i> universitários e tabelas classificatórias devem se esforçar para obter maior transparência e interoperabilidade entre diferentes sistemas de medição. Alguns <i>rankings</i> universitários têm adotado medidas louváveis para serem mais abertos sobre a escolha de indicadores e os pesos a eles atribuídos, mas outros <i>rankings</i> ainda permanecem em uma “caixa-preta”.	Fornecedores de dados, analistas e produtores de <i>rankings</i> universitários e tabelas classificatórias
	8. Os editores devem reduzir a ênfase nos fatores de impacto de periódicos como uma ferramenta de promoção e usá-los apenas no contexto de uma variedade de métricas baseadas em periódicos que fornecem uma visão mais rica do desempenho. Conforme sugerido pela DORA, esse conjunto mais amplo de indicadores pode incluir Fator de Impacto de cinco anos, EigenFactor, SCImago, tempos de edição e de publicação. Os editores, com a ajuda do Committee on Publication Ethics (COPE), devem incentivar práticas de autoria responsáveis e o fornecimento de informações mais detalhadas sobre as contribuições específicas de cada autor. Os editores também devem disponibilizar uma variedade de métricas no nível do artigo para incentivar uma mudança na direção da avaliação com base na qualidade acadêmica de um artigo, e não nos Fatores de Impacto.	Editores

THE METRIC TIDE – RECOMENDAÇÕES		
OBJETIVOS	RECOMENDAÇÕES	DESTINATÁRIOS
Melhorar a infraestrutura de dados que suporta o gerenciamento de informações de pesquisa	9. É necessário maior transparência e abertura na infraestrutura de dados de pesquisa. Um conjunto de princípios deve ser desenvolvido para tecnologias, práticas e culturas que possam apoiar o gerenciamento de informações de pesquisa abertas e confiáveis. Esses princípios devem ser adotados por financiadores, provedores de dados, administradores e pesquisadores como base para trabalhos futuros.	Órgãos de fomento de IES, conselhos de pesquisa, provedores de dados, gestores, administradores
	10. O sistema de pesquisa nacional deve aproveitar o máximo do identificador ORCID como seu sistema preferido de identificadores únicos. Os IDs do ORCID devem ser obrigatórios para todos os pesquisadores.	IES, órgãos de financiamento, gestores
	11. Identificadores também são necessários para as instituições. O uso de identificadores persistentes para instituições deve ser mais amplamente empregado em processos internos de gestão de IES e financiadores.	Órgãos de financiamento, IES, financiadores, editores
	12. Os editores devem exigir os ORCID iDs e identificadores de instituições e referência aos financiadores durante o processo de submissão de artigos, garantindo esses metadados durante todo o ciclo de vida da publicação. Isso facilitará o intercâmbio de informações sobre as atividades de pesquisa e ajudará a fornecer dados e métricas com carga mínima para pesquisadores e administradores.	Editores e provedores de dados
	13. O uso de Identificadores de Objetos Digitais (DOIs) deve ser estendido para abranger todos os resultados da pesquisa. Os DOIs também devem ser mais amplamente adotados nos processos internos das IES e dos financiadores de pesquisas. Os DOIs já predominam na esfera de publicação de periódicos – eles devem ser estendidos para cobrir outros formatos para os quais não existam um sistema identificador, como capítulos de livros e conjuntos de dados.	Órgãos nacionais de financiamento, IES
	14. É necessário mais investimento na infraestrutura de informações de pesquisa. Financiadores devem explorar oportunidades para investimentos estratégicos adicionais, principalmente para melhorar a interoperabilidade dos sistemas de gerenciamento de pesquisa.	Financiadores, IES

THE METRIC TIDE – RECOMENDAÇÕES		
OBJETIVOS	RECOMENDAÇÕES	DESTINATÁRIOS
Aumentar a utilidade dos dados e das fontes de informação existentes	15. Conselhos de financiamento do ensino superior, os financiadores e as IES devem explorar como aproveitar os dados mantidos nas plataformas existentes para apoiar os processos de avaliação da pesquisa e do ensino superior, e vice-versa. Também é necessário um debate adicional sobre os méritos das coletas nas IES e a coleta de dados em nível nacional.	Órgãos de finanças públicas, departamentos de inovação, conselhos de pesquisa, órgãos nacionais de financiamento, associações de gestores de pesquisa
	16 Departamentos de inovação devem identificar maneiras de vincular os dados coletados das plataformas relacionadas à pesquisa que sejam mais diretamente relacionados aos processos dos departamentos de inovação e dos outros departamentos, especialmente em relação à previsão, à exploração de horizonte e às prioridades de pesquisa.	Departamentos de inovação, outros departamentos governamentais, órgãos de financiamento do ensino superior
Usar métricas nos próximos ciclos avaliativos	17. Para o próximo ciclo de avaliação do REF, são feitas algumas recomendações específicas ao Higher Education Funding Council for England (HEFCE) e aos outros organismos de financiamento.	Órgãos de financiamento do ensino superior
Coordenar atividades e construir evidências	18 A comunidade de pesquisa precisa de um mecanismo para levar adiante a agenda estabelecida neste relatório. Propomos o estabelecimento de um Fórum de Métricas Responsáveis, que reuniria financiadores de pesquisa, IES e seus órgãos representativos, editores, provedores de dados e outros envolvidos para trabalhar em questões de padrões de dados, interoperabilidade, abertura e transparência. Órgãos de financiamento devem coordenar esse fórum, contando com o apoio e a experiência de outros financiadores e órgãos do setor, conforme apropriado.	Órgãos de financiamento do ensino superior, associações de gestores e administradores de pesquisa

THE METRIC TIDE – RECOMENDAÇÕES		
OBJETIVOS	RECOMENDAÇÕES	DESTINATÁRIOS
Coordenar atividades e construir evidências	19. Os financiadores de pesquisas precisam aumentar o investimento na política da ciência. É necessário maior pesquisa e inovação nessa área, para desenvolver e aplicar <i>insights</i> de computação, estatística, ciências sociais e economia para entender melhor a relação entre pesquisa, suas qualidades e seus impactos mais amplos.	Organismos de financiamento da pesquisa
	20. Um aspecto positivo deste relatório foi o debate que o documento gerou. Como uma iniciativa legada, o grupo diretor criou um <i>blog</i> (www.ResponsibleMetrics.org) como um fórum para discussão contínua das questões levantadas por este relatório. O <i>site</i> celebra práticas responsáveis, mas também apresenta as más práticas quando elas ocorrerem. Os pesquisadores são incentivados a enviar tanto exemplos de melhores práticas quanto de exemplos mais flagrantemente de usos inadequados de indicadores quantitativos no gerenciamento de pesquisas.	Comitê de acompanhamento

Fonte: adaptado de Wilsdon *et al.* (2015).

Essas recomendações são sustentadas pela noção de “métricas responsáveis” como uma maneira de estruturar os usos apropriados de indicadores quantitativos na governança, gestão e avaliação da pesquisa. Na perspectiva apresentada pelo *The metric tide*, métricas responsáveis podem ser entendidas nos termos das seguintes dimensões:

- **Robustez:** basear as métricas nos melhores dados possíveis em termos de precisão e escopo;
- **Humildade:** reconhecer que a avaliação quantitativa deve apoiar – mas não suplantiar – a avaliação qualitativa e especializada;
- **Transparência:** manter a coleta de dados e os processos analíticos abertos e transparentes, para que aqueles que estão sendo avaliados possam testar e verificar os resultados;
- **Diversidade:** considerar a variação por campo e empregar uma série de indicadores para refletir e apoiar uma pluralidade de carreiras de pesquisadores e em todo o sistema;
- **Reflexividade:** reconhecer e antecipar os efeitos potenciais e sistêmicos dos indicadores e atualizá-los sempre considerando esses resultados.

As recomendações e principais conclusões do *The metric tide* fornecem contribuição importante para avançar na discussão e na implementação de processos avaliativos centrados no uso responsável de métricas. Esse relatório, desenvolvido em parte para informar o REF usado para avaliar as universidades de todo o Reino Unido, tem sido utilizado por essas universidades também como motivação, para encabeçar um movimento que defende que o uso transparente e responsável de indicadores que deve ser empregado tanto nas avaliações internas, conduzidas dentro universidades, quanto nas avaliações externas, conduzidas para a produção de *rankings* globais de universidades.

Princípios de Hong Kong

Nesse cenário do uso responsável de métricas e indicadores, um dos mais recentes posicionamentos coletivos foi a publicação, em 2019, dos Hong Kong Principles (HKPs), que foram desenvolvidos como parte da 6th World Conference on Research Integrity, realizada em Hong Kong no período de 2 a 5 de junho de 2019, que teve como foco a integridade⁹ na pesquisa, buscando valorizar pesquisadores que se comprometem com práticas robustas, rigorosas e transparentes que impulsionam a melhoria da pesquisa, garantindo que sejam devidamente reconhecidos e recompensados por adotarem práticas e comportamentos que levam a pesquisas confiáveis.

Os HKPs foram desenvolvidos considerando que sua implementação poderia auxiliar tanto na avaliação, quanto no avanço dos pesquisadores em suas carreiras, com o objetivo de fortalecer a integridade da pesquisa. Para tanto, os HKPs apresentam cinco princípios: 1) práticas de pesquisa responsáveis; 2) relatórios transparentes; 3) ciência aberta (pesquisa aberta); 4) valorização da diversidade de tipos de pesquisa; e 5) reconhecimento da variedade de contribuições à pesquisa e à atividade acadêmica.

O Quadro 4 apresenta cada princípio, assim como a justificativa para sua consideração nos modelos de avaliação de grupos de pesquisadores ou indivíduos.

9 Embora existam alguns mecanismos para apoiar os pesquisadores a garantir a transparência em todos os estágios do projeto, da execução e da geração de relatórios, não há adoção generalizada dessas práticas em todas as disciplinas. Um fator que contribui para isso é que pouca ênfase é colocada no rigor da pesquisa ao contratar, avaliar e promover pesquisadores (Moher *et al.*, 2020).

Quadro 4 – Princípios de Hong Kong

PRINCÍPIOS DE HONG KONG	
1	Princípio 1: Avaliar os pesquisadores sobre práticas responsáveis desde a concepção até a publicação, incluindo o desenvolvimento da ideia de pesquisa, desenho da pesquisa, metodologia, execução e disseminação eficaz. <i>Fundamentação</i> O número de publicações, citações e o volume total de concessão bolsas ainda são frequentemente os indicadores predominantemente empregados pelas instituições de pesquisa para avaliar e recompensar seus pesquisadores. O fornecimento de bônus aos acadêmicos para publicação em determinados periódicos – isto é, pagamento por mérito – também é comum em muitas partes do mundo. Esses critérios de avaliação dizem pouco aos avaliadores sobre os pesquisadores e o rigor de seu trabalho; portanto, eles não são métricas particularmente “responsáveis”. Essas métricas também podem ser indevidamente influenciadas por práticas de campo e citação e fornecem pouca informação sobre as contribuições de uma publicação – e, portanto, de um pesquisador – para pesquisa e sociedade. Outros critérios são necessários para fornecer uma visão mais ampla dos indicadores das melhores práticas: por exemplo, até que ponto um pesquisador desenvolve problemas de pesquisa com o envolvimento de membros da sociedade.
2	Princípio 2: Valorize os relatórios precisos e transparentes de todas as pesquisas, independentemente dos resultados. <i>Fundamentação</i> A não publicação de todas as descobertas de todos os estudos distorce seriamente a base de evidências para a tomada de decisão. Por exemplo, uma revisão sistemática de estudos de reboxetina para o tratamento da depressão descobriu que quase três quartos dos pacientes incluídos estavam em estudos não publicados. A publicação seletiva de pesquisas com resultados positivos – isto é, vies de publicação – distorce a base de evidências da ciência e foi demonstrada em uma variedade de disciplinas, incluindo economia, psicologia e pesquisa clínica e pré-clínica em saúde. Além disso, a frequência de outros vieses de relatórios – por exemplo, resultados primários alternados sem divulgação e rotação – é de cerca de 30%. Isso é inaceitavelmente alto e diminui a confiabilidade e a integridade da pesquisa. Os Comitês de Avaliação e Promoção geralmente não dão importância suficiente ao registro de protocolos e planos de análise de dados, à publicação completa de estudos concluídos ou à disponibilização de dados, códigos e materiais.
3	Princípio 3: Valorizar as práticas da ciência aberta (pesquisa aberta) – como métodos, materiais e dados abertos. <i>Fundamentação</i> A abertura na pesquisa é mais do que apenas o acesso à produção – ela traz igualdade ao processo de pesquisa. Ela abrange uma série de práticas em todo o ciclo de vida desta. O acesso à pesquisa não deve ser sobre quem tem os recursos para pagar o alcance às pesquisas (<i>paywall</i>), normalmente os periódicos de assinatura. As decisões em saúde e políticas sociais devem ser tomadas com base no acesso a todo conhecimento de pesquisa e não apenas a uma parte dele. Uma quantidade considerável de fundos públicos é usada para pesquisa, e seus resultados podem ter um impacto social profundo. Os cientistas pré-clínicos estão se comprometendo a compartilhar abertamente seus cadernos de laboratório para otimizar a pesquisa, promover colaborações e reduzir a duplicação desnecessária. Em um esforço para dissuadir práticas questionáveis de autoria, os consórcios que avançam os padrões em informações de administração de pesquisa apoiam a taxonomia CRediT como uma maneira de os autores da pesquisa descreverem mais abertamente como cada pessoa contribuiu para um projeto de pesquisa.

PRINCÍPIOS DE HONG KONG	
3	O compartilhamento de dados é outro exemplo de abertura, mas não é uma prática comum em pesquisas clínicas – com algumas exceções, como a genética –, embora os pacientes pareçam favoráveis ao compartilhamento de seus dados, pelo menos em ensaios randomizados nos quais eles participaram. O compartilhamento de dados também não é considerado padrão em muitas outras disciplinas. Sem o compartilhamento de dados, é difícil verificar a seletividade dos relatórios; o compartilhamento de dados é essencial para enfrentar a crise de reprodutibilidade e criar confiança. Existem estimativas variáveis de qual proporção de pesquisa é disponibilizada por meios de acesso aberto, como periódicos específicos, repositórios ou <i>preprints</i>, mas está longe de 100%.
4	Princípio 4: Valorizar uma ampla gama de pesquisas e bolsas de estudos, como replicação, inovação, tradução, síntese e metapesquisa. <i>Fundamentação</i> Um sistema busca beneficiar a sociedade e incentivar a pesquisa confiável é importante e precisa levar em consideração os diferentes tipos de pesquisa: criação de novas ideias; testes; replicação das principais descobertas; síntese de pesquisas existentes; desenvolvimento e validação de novas ferramentas; medidas ou métodos etc. Diferentes indicadores e critérios precisam ser desenvolvidos e para esses diferentes tipos e estágios de pesquisa. Isso inclui diferentes prazos de avaliação para diferentes tipos de pesquisa.
5	Princípio 5: Avalie uma série de outras contribuições à pesquisa responsável e à atividade acadêmica, como revisão por pares de subsídios e publicações, orientação, divulgação e troca de conhecimentos. <i>Fundamentação</i> Conforme discutido ao lado do Princípio 1, as avaliações de pesquisa frequentemente se concentram em uma gama estreita de métricas fáceis de medir, incluindo publicações, citações e receita de financiamento. Para que o ecossistema de pesquisa funcione de maneira ideal, outras atividades de pesquisa também são essenciais. A revisão por pares continua sendo a pedra angular da avaliação da qualidade de bolsas concedidas, publicações e conferências. A qualidade das contribuições da revisão por pares para os periódicos e financiadores também deve fazer parte das avaliações de promoção e posse, assim como as contribuições para várias infraestruturas, supervisão ou regulamentos de pesquisa. Da mesma forma, as contribuições para melhorias que vão além de uma abordagem centrada no indivíduo para avaliação devem ser consideradas. Atualmente, essas atividades estão em falta nos comitês de avaliação e promoção. As contribuições para o desenvolvimento de carreiras de outras pessoas em todas as etapas de sua carreira são críticas, assim como as contribuições de vários comitês relacionados à pesquisa – por exemplo, assumindo o papel de editor científico. A melhor maneira de fazer isso sem criar mais barreiras e burocracias, no entanto, é debatida há muito tempo. Qualquer sistema de recompensa que tenha em mente toda a empresa de pesquisa e que vise promover um clima propício à pesquisa confiável e útil, com o mais alto respeito à integridade, precisa encontrar maneiras de incorporar essas funções vitais em sua estrutura geral de avaliação.

Fonte: adaptado de Moher e demais autores (2020).

Os HKPs centram-se na promoção de práticas de avaliação que buscam valorizar a integridade da pesquisa, concentrando-se principalmente no que as instituições de pesquisa podem fazer para modificar os critérios usados pelos comitês de seleção e promoção nas avaliações de suas carreiras.

Os cinco princípios formulados visam como as instituições de pesquisa buscam incentivar, recompensar e avaliar pesquisadores individuais em suas respectivas organizações.

Os HKPs enviam uma mensagem clara de que comportamentos que promovem a integridade da pesquisa precisam ser reconhecidos e recompensados. Os cinco princípios formulados pelos pesquisadores responsáveis pelo desenvolvimento dos HKPs chamam atenção para o fato de que as instituições devem incentivar, recompensar e avaliar investigadores individuais por comportamentos que promovem a integridade da pesquisa em suas respectivas organizações. Esses temas precisam ser considerados com atenção durante uma avaliação de um grupo de pesquisadores no momento de tomar decisões sobre alocações de fundos ou políticas de recursos humanos. Além disso, essas questões dizem respeito à justiça social e à relevância social da pesquisa, e não à integridade da pesquisa (Moher *et al.* 2020).

De acordo com Moher e demais autores (2020), se implementados, os HKPs podem desempenhar um papel crítico nas avaliações baseadas em evidências que colocam o rigor da pesquisa no centro da avaliação, além de abrir a pesquisa para o benefício mais amplo da sociedade.

Declaração de Princípios Folec-Clacso

No contexto da repensar os métodos e critérios de avaliação da pesquisa científica na América Latina e no Caribe, considerando as necessidade e os desafios locais, em 2022 o Conselho Latino-Americano de Ciências Sociais (Clacso) e o Foro Latinoamericano de Evaluación Científica (Folec) uniram esforços para elaborar e aprovar um posicionamento que representa uma resposta coletiva à necessidade de promover uma avaliação mais justa, equitativa e contextualizada da pesquisa científica na região, que busca incentivar uma ciência que produza conhecimento de alta qualidade, e também contribua para a solução de problemas sociais, econômicos e ambientais enfrentados pela comunidade latino-americana e caribenha. O resultado desse esforço está documentado na *Declaração de princípios: uma nova avaliação acadêmica e científica para uma ciência com relevância social na América Latina e no Caribe* (Conselho Latino-Americano de Ciências Sociais, 2022).

A seguir, são apresentadas, de forma resumida, as principais práticas de avaliação propostas na Declaração de Princípios Folec-Clacso para a construção de um novo sistema de avaliação acadêmica e científica da América Latina e do Caribe mais alinhado com as necessidades e desafios da região.

Sobre os objetivos da avaliação

1. O principal objetivo da avaliação científica e acadêmica é garantir o desenvolvimento de uma ciência de qualidade, com relevância social e ética, respeitando os Direitos Humanos e comprometida com a construção de sociedades justas, democráticas e igualitárias.
2. É necessária uma adaptação ao atual estágio da ciência aberta, através de novas políticas de avaliação que dêem prioridade à avaliação qualitativa da pesquisa [...].
3. O conhecimento científico é uma construção coletiva, pelo que é fundamental que a avaliação acadêmica e científica pondere adequadamente o trabalho em equipe e as suas diferentes formas de organização e construção.

Sobre os processos de avaliação

1. É fundamental recuperar o controle da comunidade acadêmica e científica sobre os processos de avaliação e os seus indicadores [...].
2. Os indicadores da produção publicada a serem utilizados nos processos de avaliação devem incluir aqueles produzidos pelos serviços de indexação regional e internacional de periódicos de qualidade (Latindex Catálogo, Redalyc, SciELO, DOAJ, entre outros), bem como índices nacionais de periódicos de qualidade para neutralizar os indicadores WoS e Scopus. [...].

3. A noção de “impacto” da pesquisa científica deve ser expandida para incluir a “relevância social” do conhecimento [...].
4. É fundamental reconhecer, nos processos de investigação colaborativos e participativos, o contributo do conhecimento fornecido por atores sociais fora do campo académico ligados aos temas sob investigação [...].
5. [...] O multilinguismo favorece o desenvolvimento de pesquisa socialmente relevante e contribui para sustentar a diversidade cultural. [...].
6. Os processos de avaliação devem ser evolutivos, autorreflexivos, transparentes e participativos, promovendo mecanismos que estimulem o diálogo e a aprendizagem mútua e que garantam melhorias contínuas [...].
7. Considerar a revisão por pares como parte das atividades de quem pesquisa e como uma contribuição relevante para a comunidade científica e académica [...].
8. É essencial garantir a representação equitativa das mulheres e das diversidades nos órgãos e processos de avaliação com um mínimo de paridade e nas prioridades de investigação e seus temas [...].
9. Abordar nas fases iniciais das carreiras académicas e científicas os problemas de inclusão que resultam de práticas de avaliação inadequadas [...].

Sobre sistemas de informação e indicadores

1. Os sistemas de informação das organizações públicas de ciência e tecnologia, agências de fomento à pesquisa e universidades devem refletir a trajetória das pessoas que realizam ensino, pesquisa, extensão, ligação e intervenção social e daqueles que estão em formação, [...] respeitando a diversidade de estruturas institucionais e disciplinares, culturas e seus diversos formatos de comunicação.

2. Os indicadores de citações extraídos de bases de dados limitadas no seu âmbito geográfico, linguístico e disciplinar não devem ser considerados uma medida válida para fazer comparações da produção científica entre indivíduos, instituições e países [...] (Conselho Latino-Americano de Ciências Sociais, 2022, p. 3-6, tradução nossa).

Ao propor diretrizes e recomendações para uma avaliação mais contextualizada e responsável da pesquisa, a Declaração de Princípios Folec-Clasco representa um marco importante para a comunidade acadêmica latino-americana e caribenha, que com o engajamento de diversos atores na implementação das recomendações do documento tem o potencial de transformar significativamente a cultura de avaliação da pesquisa na região, promovendo também práticas mais justas, transparentes e socialmente responsáveis de avaliação da pesquisa científica.

Ranking de universidades: implicações do movimento pelo uso transparente e responsável de indicadores

O atual sistema de avaliação da pesquisa está sob pressão. Nas últimas duas décadas, o investimento global em Pesquisa e Desenvolvimento (P&D) triplicou, atingindo cerca de 2 bilhões de dólares por ano (Organização das Nações Unidas para a Educação, a Ciência e a Cultura, 2021). Esse aumento de investimento em pesquisa traz consigo uma cultura de responsabilidade (*accountability*) que coloca pressão sobre os indivíduos e as instituições envolvidas em pesquisa, gerando, muitas vezes, anomalias ou incentivos distorcidos. Além disso, essa cultura também levou a aspirações elevadas, tais como: manter a qualidade e reduzir o desperdício, minimizar erros e ineficiências da pesquisa; maximizar a inclusão e a diversidade; potencializar a pesquisa como um bem público global; e incentivar empreendimentos científicos mais abertos e engajados (Rijcke *et al.*, 2023).

Como temos observado, em várias partes do mundo, há um aumento do reconhecimento de que um conjunto estreito e simplista de métricas e indicadores de avaliação não captura satisfatoriamente a qualidade, a utilidade, a integridade e a diversidade da pesquisa. Campanhas para limitar uso

inadequado de métricas, ampliar os critérios de qualidade e transformar a cultura de pesquisa de forma mais sistemática, por meio de manifestos, declarações, princípios e reformas prepararam o cenário para um debate global sobre a necessidade de reformar a avaliação da pesquisa (Conselho Latino-Americano de Ciências Sociais, 2022; Declaration on Research Assessment, 2012; Hicks *et al.*, 2015; Moher *et al.*, 2020; Wilsdon *et al.*, 2015). Essas vozes agora pedem uma mudança de manifestos para ação (Rijcke *et al.*, 2023). Isso ocorre em meio a mudanças transformacionais na forma como a pesquisa é realizada e comunicada, como, por exemplo, o surgimento de estruturas de pesquisa abertas e de mídias sociais, a mudança para ciência orientada por missões e de caráter transdisciplinar, o crescimento da revisão por pares aberta e o potencial transformador da inteligência artificial e o aprendizado de máquina exigem novas reflexões sobre como a pesquisa e os pesquisadores são avaliados (Rijcke *et al.*, 2023).

Assim, aqueles envolvidos com funções de gestão da pesquisa há muito tempo enfrentam um dilema ao construir equipes eficientes, realizar uma avaliação justa, responsável, ao mesmo tempo que mantêm um “olho” no reconhecimento externo representado principalmente pelos *rankings* de universidades, cuja trajetória se dá na direção oposta da que tem sido defendida pela comunidade acadêmica.

Nesse cenário efervescente, era de se esperar que em algum momento a onda do movimento pelo uso transparente e responsável de indicadores na avaliação, potencializado também pelos avanços recentes do movimento em defesa da abertura, integridade e transparência do processo de pesquisa promovidos pela ciência aberta, alcançasse os *rankings* de universidades.

Rankings de universidades: “somos mais do que nossa classificação”

Apesar do acalorado debate sobre o uso responsável de métrica e indicadores, e das diversas críticas da comunidade acadêmica aos *rankings* globais de universidades em mais de duas décadas, as instituições responsáveis pela elaboração dessas classificações parecem permanecer incólumes e continuam a proliferar, enquanto os usuários finais continuam a confiar nas

classificações, ainda que sejam consideradas um exercício “preguiçoso” de avaliação da qualidade das universidades (Gadd, 2020).

Nesse contexto, repensar e exigir que também os *rankings* de universidades sejam produzidos e utilizados de forma justa e responsável que a rede International Network of Research Management Societies¹⁰ (Inorms), uma organização global que reúne sociedades profissionais envolvidas na gestão de pesquisa em todo o mundo, desenvolveu instrumentos para orientar a mudança organizacional. Com o apoio de seu grupo Research Evaluation Group, a Inorms desenvolveu o SCOPE¹¹ *framework for research evaluation*, que começa definindo o que é valorizado, quem está sendo avaliado e por quê. O instrumento orienta uma série de instituições internacionais relacionadas à gestão de pesquisa a trabalharem em prol de uma avaliação de pesquisa melhor, mais justa e mais significativa (Inorms Research Evaluation Group, 2001, 2021; Rijcke *et al.*, 2023).

Desde 2023, a Inorms também lançou a campanha *More Than Our Rank*, liderada por acadêmicos, pesquisadores e ativistas preocupados com as limitações e consequências negativas dos *rankings* universitários globais. A *More Than Our Rank* busca aumentar a conscientização sobre as limitações dos *rankings* universitários e, ao mesmo tempo, promover uma discussão mais ampla sobre como avaliar a qualidade e o impacto das IES. Dentre as preocupações levantadas pela iniciativa estão: a simplificação da avaliação da qualidade universitária, a exclusão de aspectos importantes da excelência acadêmica e a pressão indesejada que os *rankings* podem exercer sobre as universidades e os pesquisadores. A campanha incentiva uma abordagem mais holística e contextualizada para avaliar as universidades, levando em consideração uma variedade de indicadores qualitativos e quantitativos e

10 A Inorms é uma organização global que reúne sociedades profissionais envolvidas na gestão de pesquisa em todo o mundo. Fundada em 2001, segundo a Inorms, seu objetivo é promover a cooperação e o intercâmbio de boas práticas entre profissionais de gestão de pesquisa, fornecendo uma plataforma para compartilhar conhecimento, desenvolver competências e abordar desafios comuns enfrentados por aqueles que trabalham nessa área.

11 O SCOPE foi desenvolvido pelo Research Evaluation Group (REG) da Inorms como modelo de cinco etapas para avaliar a pesquisa de forma mais aberta e responsável. É um guia prático com passo a passo concebido para auxiliar os gestores de pesquisa, ou qualquer pessoa envolvida na condução de avaliações de pesquisa, a conceber e planejar novas avaliações, bem como a revisar as avaliações existentes (Inorms Research Evaluation Group, 2021, 2024).

reconhecendo a diversidade de missões e objetivos das IES (Inorms Research Evaluation Group, 2001).

De acordo com os responsáveis pela campanha, trata-se de uma iniciativa para instituições que se orgulham da sua posição no *ranking*, mas também reconhecem as limitações dos indicadores utilizados, e representa também aqueles que sentem que os *rankings* não refletem os seus pontos fortes ou missão institucional. Cada instituição no mundo é muito mais do que a apenas sua classificação em um *ranking*, e a iniciativa oferece a essas instituições a oportunidade de se posicionar e explicar por que são mais do que sua sua posição alcançada em qualquer *ranking* nacional ou global (Inorms Research Evaluation Group, 2001).

Considerações finais

As atividades acadêmico-científicas são financiadas, na maioria dos países do mundo, total ou parcialmente por recursos públicos. Mesmo universidades privadas são, pelo menos em parte, financiadas com recursos públicos. Sendo assim, nosso trabalho se beneficia de financiamento público e, por razões fundamentais, deve-se realizar a prestação de contas à sociedade. A partir da academia, podemos e devemos discutir as formas e os métodos empregados na avaliação de projetos e carreiras acadêmicas, mas não a sua necessidade.

Um aumento na avaliação externa leva previsivelmente a um aumento na avaliação interna (Gadd, 2019). Avaliações como as conduzidas pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) têm desempenhado papel importante no desenvolvimento da pós-graduação brasileira, mas, como qualquer processo avaliativo de longa duração, devem ser periodicamente repensadas para o aprimoramento dos mecanismos de atribuição de valor e diferenciação entre objetos avaliados (Iriart *et al.*, 2015). Essa revisão é necessária, sobretudo, para que se evite a cristalização de parâmetros e critérios e para que se discutam criticamente as consequências que produzem, evitando distorções (Franceschet; Costantin, 2011; Moed, 2008).

A avaliação tem um papel indutor importante no desenvolvimento de uma área, uma vez que os critérios adotados orientam os caminhos para a organização dos programas de pesquisa e apontam tendências. Portanto, se desejamos incentivar uma maior adesão ao acesso aberto e à ciência aberta, melhorar a reprodutibilidade ou o rigor da pesquisa, aperfeiçoar a igualdade e a diversidade na academia, ou pressionar por mudanças na forma como os *rankings* reconhecem e avaliam o desempenho, é necessário alterar a maneira como a pesquisa é avaliada, de modo a reconhecer e recompensar a adoção dessas práticas pelos pesquisadores. O método de avaliação tem um efeito significativo em todo o esforço acadêmico. Isso se chama Lei de Campbell, muitas vezes resumida como a maneira de medir alguém afeta a maneira como eles se comportam (Campbell, 1979) e esse tipo de efeito pode afetar avaliação da pesquisa em seus mais distintos níveis, incluindo o individual, os grupos de pesquisa, os programas de pós-graduação, as disciplinas, as áreas do conhecimento, as instituições, as universidades e também a produção científica em nível nacional.

Como destaca Gadd (2019), identificar o problema é muito mais simples do que resolvê-lo – especialmente quando está tão profundamente enraizado nas culturas acadêmicas globais existentes. Os compromissos, posicionamentos coletivos e manifestos aqui apresentados destacam a necessidade de mudar a forma como estamos “medindo” e avaliando a pesquisa, e nos convidam a empregar as melhores práticas de avaliação para coibir abusos e promover o uso responsável de métricas e indicadores nos processos avaliativos, alcançando também o nível macro, aqui representado pelos *rankings* globais de universidades.

Referências

- ALTBACH, P. G. The globalization of college and university rankings. *Change: The Magazine of Higher Learning*, [Washington], v. 44, n. 1, p. 26-31, 2012.
- AMSLER, S. S.; BOLSMANN, C. University ranking as social exclusion. *British Journal of Sociology of Education*, Oxford, v. 33, n. 2, p. 283-301, 2012. DOI 10.1080/01425692.2012.649835. Disponível em: <https://bpbeu-w2.wpmucdn.com/blogs.lincoln.ac.uk/dist/a/1098/files/2013/01/>

Amsler_Bolsmann_University_ranking_as_social_exclusion_final_20121.pdf. Acesso em: 11 out. 2024.

BALL, S. J. Educação Global S.a.: novas redes políticas e o imaginário neoliberal. Ponta Grossa: UEPG, 2014.

BATY, P. Ranking confession. *Inside Higher Ed*, [s. l.], 14 Mar. 2010. Disponível em: <https://www.insidehighered.com/views/2010/03/15/ranking-confession>. Acesso em: 11 out. 2024.

BOUGNOL, M.-L.; DULÁ, J. H. Technical pitfalls in university rankings. *Higher Education*, [s. l.], v. 69, p. 859-866, 2015.

BRANKOVIC, J.; RINGEL, L.; WERRON, T. How rankings produce competition: The case of global university rankings. *Zeitschrift für Soziologie*, Stuttgart, v. 47, n. 4, p. 270-288, 2018.

BRINK, C. *The soul of a university: Why excellence is not enough*. [s. l.]: Bristol University Press, 2018.

CALDERÓN, A. I.; FRANÇA, C. M.; GONÇALVES, A. Tendências dos rankings acadêmicos de abrangência nacional de países do espaço ibero-americano: os rankings dos jornais El Mundo, El Mercurio, Folha de São Paulo, Reforma e El Universal. *EccoS Revista Científica*, São Paulo, n. 44, p. 117-141, set./dez. 2017. DOI 10.5585/EccoS.n44.7943. Disponível em: <https://periodicos.uninove.br/eccos/article/view/7943/3647>. Acesso em: 9 out. 2024.

CAMPBELL, D. T. Assessing the impact of planned social change. *Evaluation and Program Planning*, New York, v. 2, n. 1, p. 67-90, 1979. DOI 10.1016/0149-7189(79)90048-X. Disponível em: https://www.sciencedirect.com/science/article/pii/014971897990048X?casa_token=7gzgGcZTYFAAAAAA:RUk5N6yzAO9p9dCj5veOhmPiSo8kZz-uBwUOBds9ZVmb0M2hfNzSv6lnTJqtoYvhYymcpQPAxw. Acesso em: 11 out. 2024.

CHU, J. Cameras of merit or engines of inequality? College ranking systems and the enrollment of disadvantaged students. *American Journal of Sociology*, Chicago, v. 126, n. 6, p. 1307-1346, 2021.

COLLINS, F. L.; PARK, G.-S. Ranking and the multiplication of reputation: reflections from the frontier of globalizing higher education. *Higher Education*, [s. l.], v. 72, p. 115-129, 2016.

CONSELHO LATINO-AMERICANO DE CIÊNCIAS SOCIAIS.

Declaración de principios: una nueva evaluación académica y científica para una ciencia con relevancia social en América Latina y el Caribe. Buenos Aires: CLACSO, 2022. Disponível em: <https://biblioteca-repositorio.clacso.edu.ar/bitstream/CLACSO/169563/1/Declaracion-CLACSO-FOLEC-version-extendida.pdf>. Acesso em: 11 out. 2024.

DECLARATION ON RESEARCH ASSESSMENT. *San Francisco*

Declaration of Research Assessment (DORA). San Francisco, 2012. Disponível em: <https://sfdora.org/read/>. Acesso em: 28 Mar. 2024.

DILL, D. D. Convergence and diversity: The role and influence of university rankings. In: KEHM, B. M.; STENSAKER, B.; STENSAKER, B. (ed.). *University rankings, diversity, and the new landscape of higher education*. [s. l.: s. n.], 2006. p. 97-116.

DILL, D. D.; SOO, M. Academic quality, league tables, and public policy: A cross-national analysis of university ranking systems. *Higher Education*, [s. l.], v. 49, p. 495-533, 2005.

ERKKILÄ, T.; PIIRONEN, O. Rankings and global knowledge governance. In: HAZELKORN, E.; MIHUT, G. (ed.). *Research Handbook on University Rankings: Theory, Methodology, Influence and Impact*. Cheltenham: Edward Elgar Publishing, 2021. p. 54-66.

ESPELAND, W. N.; SAUDER, M. Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, Chicago, v. 113, n. 1, p. 1-40, July. 2007. Disponível em: <https://sociology.northwestern.edu/documents/faculty-docs/faculty-research-article/rankings-and-reactivity-2007.pdf>. Acesso em: 11 out. 2024.

FERNÁNDEZ-CANO, A. *et al.* Questioning the Shanghai Ranking methodology as a tool for the evaluation of universities: An integrative review. *Scientometrics*, [Budapest], v. 116, p. 2069-2083, 2018. DOI 10.1007/s11192-018-2814-7. Disponível em: <https://link.springer.com/article/10.1007/s11192-018-2814-7>. Acesso em: 11 out. 2024.

FRANCESCHET, M.; COSTANTINI, A. The first Italian research assessment exercise: A bibliometric perspective. *Journal of Informetrics*, [s. l.], v. 5, n. 2, p. 275-291, Apr. 2011.

- GADD, E. Influencing the changing world of research evaluation. *Insights*, [s. l.], v. 32, n. 1, p. 6, 2019. DOI 10.1629/uksg.491. Disponível em: <https://insights.uksg.org/articles/456/files/submission/proof/456-1-4368-2-10-20190926.pdf>. Acesso em: 11 out. 2024.
- GADD, E. University rankings need a rethink. *Nature*, [s. l.], v. 587, n. 7835, p. 523-524, Nov. 2020. Disponível em: <https://www.nature.com/articles/d41586-020-03312-2>. Acesso em: 11 out. 2024.
- GAPPER, J. University rankings are just an educated guess. *Financial Times*, [s. l.], Sept. 2021. Disponível em: <https://www.ft.com/content/21da8a6b-d5e9-473a-86e0-056c489d55bf>. Acesso em: 28 mar. 2024.
- GINGRAS, Y. *Bibliometrics and research evaluation: Uses and abuses*. [Cambridge]: MIT Press, 2016.
- HALLONSTEN, O. Stop evaluating science: A historical-sociological argument. *Social Science Information*, London, v. 60, n. 1, p. 7-26, 2021. DOI 10.1177/0539018421992204. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/0539018421992204>. Acesso em: 11 out. 2024.
- HAMANN, J. The visible hand of research performance assessment. *Higher Education*, [s. l.], v. 72, p. 761-779, 2016.
- HAMANN, J.; RINGEL, L. The discursive resilience of university rankings. *Higher Education*, [s. l.], v. 86, n. 4, p. 845-863, 2023. DOI 10.1007/s10734-022-00990-x. Disponível em: <https://link.springer.com/article/10.1007/s10734-022-00990-x>. Acesso em: 11 out. 2024.
- HAMANN, J.; SCHMIDT-WELLENBURG, C. The double function of rankings: Consecration and dispositif in transnational academic fields. In: SCHMIDT-WELLENBURG, C.; BERNHARD, S. (ed.). *Charting Transnational Fields*. London: Routledge, 2020. p. 160-177.
- HAZELKORN, E. *Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence*. [S. l.]: Palgrave Macmillan, 2011.
- HICKS, D. et al. The Leiden Manifesto for research metrics. *Nature*, [s. l.], v. 520, p. 429-431, 2015. Disponível em: <https://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351>. Acesso em: 28 mar. 2024.

HOLMES, R. *Observations on the Indian ranking boycott*. [S. l.], 2020.

Disponível em: <https://rankingwatch.blogspot.com/2020/05/observatinss-on-indian-ranking-boycott.html>. Acesso em: 28 mar. 2024.

INORMS RESEARCH EVALUATION GROUP, [s. l.], 2001. Disponível em: <https://inorms.net/research-evaluation-group/>. Acesso em: 28 mar. 2024.

INORMS RESEARCH EVALUATION GROUP. *The SCOPE Framework: a five-stage process for evaluating research responsibly*.

[S. l.]: INORMS, 2021. Disponível em: <https://inorms.net/wp-content/uploads/2022/03/21655-scope-guide-v10.pdf>. Acesso em: 28 mar. 2024.

IOANNIDIS J. P. A.; KHOURY, M. J. Assessing value in biomedical research: the PQRS of appraisal and reward. *JAMA*, [s. l.], v. 312, n. 5, p. 483-484, Aug. 2014. DOI 10.1001/jama.2014.6932. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4687964/pdf/nihms744159.pdf>. Acesso em: 28 mar. 2024.

IREG OBSERVATORY ON ACADEMIC RANKING AND EXCELLENCE. *IREG Inventory of National Rankings*. Bruxelles, 2023.

Disponível em: <https://ireg-observatory.org/en/initiatives/ireg-inventory-of-national-rankings/>. Acesso em: 28 abr. 2023

IRIART, J. A. B. *et al.* A avaliação da produção científica nas subáreas da Saúde Coletiva: limites do atual modelo e contribuições para o debate. *Cadernos de Saúde Pública*, Rio de Janeiro, v. 31, p. 2137-2147, 2015. DOI 10.1590/0102-311X00065515. Disponível em: <https://www.scielo.br/j/csp/a/6x7HTyXCzTMWxYn7sXX6wHG/?format=pdf&lang=pt>. Acesso em: 11 out. 2024.

JOHNES, J. University rankings: What do they really show? *Scientometrics*, [Budapest], v. 115, n. 1, p. 585-606, 2018.

KAIDESOJA, T. A theoretical framework for explaining the paradox of university rankings. *Social Science Information*, London, v. 61, n. 1, p. 128-153, 2022. DOI 10.1177/05390184221079470. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/05390184221079470>. Acesso em: 11 out. 2024.

KAUPPINEN, I. *et al.* Blurring boundaries and borders: Interlocks between AAU institutions and transnational corporations. In: SLAUGHTER, S.; BARRETT, T. J. (ed.). *Higher education, stratification, and workforce*

development: Competitive advantage in Europe, the US, and Canada. [S. l.]: Springer, 2016. p. 35-57.

KEHM, B. M. Global University Rankings: Impacts and Applications. In: BIAGIOLI, M.; LIPPMAN, A. (ed.). *Gaming the metrics: misconduct and manipulation in academic research*. Cambridge: MIT Press, 2020. *E-book*. p. 93-100. DOI 10.7551/mitpress/11087.003.0009. Disponível em: <https://direct.mit.edu/books/oa-edited-volume/4598/chapter/211135/Global-University-Rankings-Impacts-and>. Acesso em: 7 out. 2024.

KROTH, A.; DANIEL, H.-D. Internationale Hochschulrankings: Ein methodenkritischer Vergleich. *Zeitschrift für Erziehungswissenschaft*, [s. l.], v. 11, n. 4, p. 542-558, 2008.

LEIBER, T. University governance and rankings: the ambivalent role of rankings for autonomy, accountability and competition. *Beiträge zur Hochschulforschung*, [s. l.], v. 39, p. 30-51, 2017. Disponível em: https://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/deliver/index/docId/37978/file/3-4-2017_Leiber.pdf. Acesso em: 11 out. 2024.

MARGINSON, S. University rankings and social science. *European Journal of Education*, Oxford, v. 49, n. 1, p. 45-59, 2014.

MOED, H. F. UK Research Assessment Exercises: Informed judgments on research quality or quantity? *Scientometrics*, [Budapest], v. 74, n. 1, p. 153-161, 2008. DOI 10.1007/s11192-008-0108-1. Disponível em: <https://link.springer.com/content/pdf/10.1007/s11192-008-0108-1.pdf>. Acesso em: 11 out. 2024.

MOHER, D. *et al.* The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLoS Biol*, [s. l.], v. 18, n. 7, p. e3000737, 2020. DOI 10.1371/journal.pbio.3000737. Disponível em: <https://journals.plos.org/plosbiology/article/file?id=10.1371/journal.pbio.3000737&type=printable>. Acesso em: 28 mar. 2024.

MOLLIS, M.; MARGINSON, S. The assessment of universities in Argentina and Australia: Between autonomy and heteronomy. *Higher Education*, [s. l.], v. 43, p. 311-330, 2002.

MÜNCH, R. *Academic capitalism: Universities in the global struggle for excellence*. London: Routledge, 2014.

MÜNCH, R. The colonization of the academic field by rankings: Restricting diversity and obstructing the progress of knowledge. In: ERLILÃ, T. (ed.). *Global university rankings: Challenges for European higher education*. London: Palgrave Macmillan UK, 2013. p. 196-219.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA. *UNESCO Science Report: The Race Against Time for Smarter Development*. Paris: UNESCO, 2021. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000377250/PDF/377250eng.pdf.multi>. Acesso em: 28 mar. 2024.

RIJCKE, S. de *et al.* *The future of research evaluation: a synthesis of current debates and developments: discussion paper*. [S. l.: s. n.], 2023. DOI 10.24948/2023.06. Disponível em: <https://www.interacademies.org/sites/default/files/2023-05/2023-05-11%2BEvaluation%2B-%2BWEB.pdf>. Acesso em: 28 mar. 2024.

ROBERTSON, S. L.; OLDS, K. *World University Rankings: On the New Arts of Governing (Quality)*. Reino Unido: Centre for Globalisation, Education and Societies, 2012.

SALMI, J.; SAROYAN, A. League tables as policy instruments: Uses and misuses. *Higher Education Management and Policy*, Paris, v. 19, n. 2, p. 1-38, 2007. Disponível em: <https://www.oecd-ilibrary.org/docserver/hemp-v19-art10-en.pdf?expires=1727083916&id=id&accname=guest&checksum=4F1F388083EA59726086EB157EBE38CA>. Acesso em: 11 out. 2024.

SANTOS, S. M. dos. *O desempenho das universidades brasileiras nos rankings internacionais: áreas de destaque da produção científica brasileira*. 2015. Tese (Doutorado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2015. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27151/tde-26052015-122043/publico/SOLANGEMARIADOSSANTOSVC.pdf>. Acesso em: 28 mar. 2024.

SANTOS, S. M. dos. Rankings internacionais de universidade: comparação e desempenho por áreas. In: MARCOVICH, J. (org.). *Repensar a universidade: desempenho acadêmico e comparações internacionais*. São Paulo: Com-Arte: Fapesp, 2018. p. 63-94. DOI 10.11606/9788571661868. Disponível em: <https://www.livrosabertos.abcd.usp.br/portaldelivrosUSP/catalog/view/224/203/937>. Acesso em: 11 out. 2024.

- SCHMOCH, U. The informative value of international university rankings: Some methodological remarks. In: WELPE, I. *et al.* (ed.). *Incentives and performance: Governance of research organizations*. Cham: Springer International Publishing, 2014. p. 141-154.
- SHORE, C.; WRIGHT, S. Audit culture revisited: Rankings, ratings, and the reassembling of society. *Current Anthropology*, Chicago, v. 56, n. 3, p. 421-444, 2015.
- SURAPPA, M. K. World university rankings and subject ranking in engineering and technology (2015-2016): A case for greater transparency. *Current Science*, [s. l.], v. 111, n. 3, p. 461-464, 2016.
- THIENGO, L. C. Universidades de Classe Mundial e o Consenso pela Excelência: Tendências e Manifestações Globais e Locais. 2018. Tese (Doutorado em Educação) – Universidade Federal de Santa Catarina, Florianópolis, 2018. Disponível em: <https://repositorio.ufsc.br/bitstream/handle/123456789/190910/PEED1316-T.pdf?sequence=-1&isAllowed=y>. Acesso em: 11 out. 2024.
- TIME to remodel the journal impact factor. *Nature*, [s. l.], v. 535, n. 7613, p. 466, 2016. Disponível em: <https://www.nature.com/news/time-to-remodel-the-journal-impact-factor-1.20332>. Acesso em: 22 mar. 2024.
- USHER, A.; SAVINO, M. A Global Survey of University Ranking and League Tables. *Higher Education in Europe*, Bucharest, v. 32, n. 1, 2007, p. 5-15.
- VAN HOUTUM, H.; VAN UDEN, A. The autoimmunity of the modern university: How its managerialism is self-harming what it claims to protect. *Organization*, [s. l.], v. 29, n. 1, p. 197-208, 2022. DOI 10.1177/1350508420975347. Disponível em: <https://journals.sagepub.com/doi/pdf/10.1177/1350508420975347>. Acesso em: 11 out. 2024.
- WEBSTER, D. S. *Academic quality rankings of American colleges and universities*. [S. l.]: Schenkman, 1986.
- WILSDON, J. *et al.* *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. [S. l.: s. n.], 2015. DOI 10.13140/RG.2.1.4929.1363. Disponível em: <https://www.ukri.org/wp-content/uploads/2021/12/RE-151221-TheMetricTideFullReport2015.pdf>. Acesso em: 11 out. 2024.

Parte II

Infraestructuras

CAPÍTULO 4

BrCris: uma infraestrutura informacional aberta sobre os ecossistemas brasileiros de pesquisa científica e inovação

Vivian dos Santos Silva

Adilson Luiz Pinto

Washington Luis R. de Carvalho Segundo

Thiago Magela Rodrigues Dias

Introdução

Sistemas de Recomendação são contribuições para auxiliar as pessoas na busca e na visualização de conteúdos focados em seu perfil, e estão disponíveis também no mundo acadêmico/científico. Representados como um filtro colaborativo de informações (Cazella *et al.*, 2009), eles podem ser um conjunto de dados que, quando combinados, criam relevância informacional mais próxima das necessidades e dos desejos de cada indivíduo. Esses Sistemas de Recomendação podem ser desenvolvidos a partir de opiniões de pessoas de uma comunidade estudada (Resnick; Varian, 1997), mas também podem ser projetados por *sites* comerciais para sugerir novos serviços ou produtos (Schafer; Konstan; Riedl, 1999), induzindo clientes ou usuários a preferirem um produto A a outro B.

Esse tipo de sistema de recomendação trabalha com um fator matemático que vincula as respostas dos usuários (R) ao combinar a identificação dos usuários e dos itens a serem avaliados, buscando uma pontuação para cada item ($R: \text{usuários} \times \text{itens} = \text{pontuação}$). Portanto, de acordo com essa abordagem, tratada principalmente no setor empresarial, chegamos a recomendações para os itens mais procurados (i) com pontuações mais significativas.

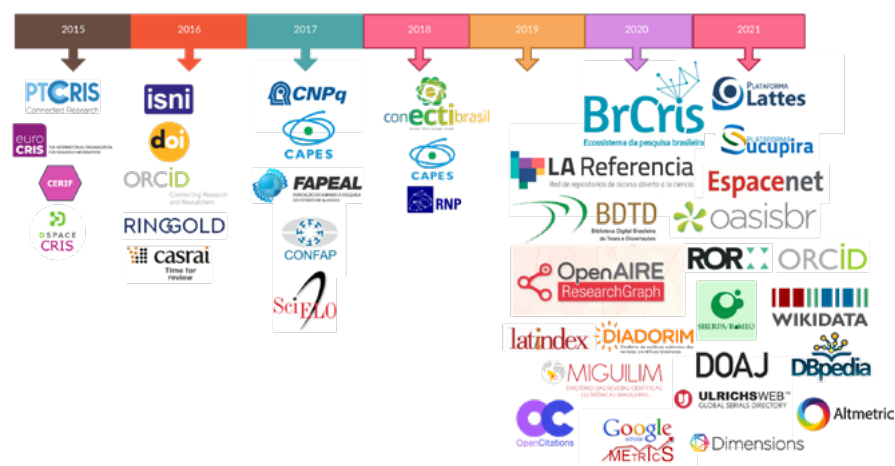
Além disso, podemos chegar a um conjunto de itens (k) em vez de apenas um. Essa recomendação é destinada ao usuário final (u) (Cazella, 2006).

$$\forall u \in users, i'_u = \underset{i \in itens}{argMax} R(u, i)$$

No contexto de sistemas de cunho científico/acadêmico, é provável que o público seja mais seletivo e especializado. Assim, suas necessidades e desejos nem sempre serão atendidos com perguntas simples e aplicação de fórmulas pensadas para o mercado. Nesse sentido, alguns modelos já estão sendo aplicados visando questões de aprendizagem (Barcellos *et al.*, 2007), usuários de unidades de informação (Krebs; Rocha; Ribeiro, 2017), e para recomendações de trabalhos científicos (Grava, 2016; Silva, 2020). Dentro do projeto Ecossistema de Informação da Pesquisa Científica Brasileira (BrCris), estão sendo desenvolvidos vários Sistemas de Recomendação científica e tecnológica, com base nas visões de: (1) um sistema de reconhecimento de periódicos acadêmicos de uma área/campo de conhecimento específico; (2) um sistema de reconhecimento de especialistas em determinadas áreas/campos para sua publicação científica; (3) um sistema de reconhecimento de especialistas em determinadas áreas/campos pela sua atividade tecnológica, ou seja, produção de patentes; e (4) um sistema de reconhecimento de especialistas em determinadas áreas/campos pela orientação de teses/dissertações e participação assídua em comissões de defesa de teses/dissertações.

Apesar de, em última instância, servir como uma base sólida para Sistemas de Recomendação, o projeto BrCris surge com o objetivo mais amplo de reunir em um único ambiente todas as informações relativas ao universo de pesquisa científica no Brasil, integrando dados de diversas fontes em uma base unificada (Pinto *et al.*, 2022). No final de 2014, guiada pelo desenvolvimento do projeto Portuguese Current Research Information System (PTCris), a equipe do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) iniciou um trabalho para estudo e adoção de padrões de descrição de recursos e interoperabilidade de sistemas de informação científica, identificadores persistentes e parcerias entre instituições do setor público (Figura 1).

Figura 1 - Histórico do projeto BrCris



Fonte: dados da pesquisa (2024).

Em 2019, foi implementado um projeto formal com recursos obtidos de uma emenda parlamentar para o mapeamento e a organização das informações científicas relacionadas à pesquisa brasileira, que se encontram dispersas em diferentes fontes abertas, nacionais e internacionais. Dentre essas fontes, destacam-se: (1) Base de Dados de Teses e Dissertações do IBICT e Portal Brasileiro de Publicações e Ddos Científicos em Acesso Aberto (Oasisbr); (2) plataformas Lattes e Sucupira; (3) plataforma LA Referencia; (4) OpenAIRE e ResearchGate; (5) Wikidata; e (6) OpenCitations. Após o mapeamento das bases, os dados foram então coletados por meio da Interface de Programação de Aplicativo (API) pública ou acesso direto às informações disponíveis nos *sites* de cada fonte. Todas as informações coletadas foram normalizadas para se adequar ao padrão de importação para a plataforma LA Referencia, na qual os dados são carregados e processados.

A proposta para a criação do BrCris se fundamenta também na experiência europeia, o euroCRIS¹, que armazena, gerencia e possibilita a troca de metadados contextuais sobre atividades de pesquisa financiadas por uma agência governamental ou conduzidas em uma organização executora de pesquisa.

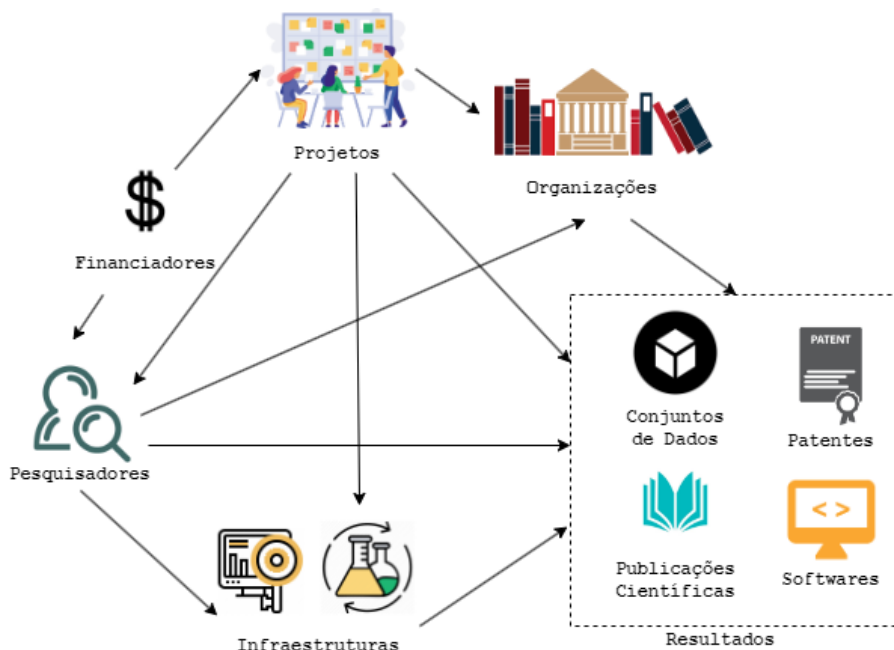
1 Disponível em: <https://www.eurocris.org/>.

Até maio de 2022, essa iniciativa contava com 330 instituições associadas², incluindo o IBICT, que é membro desde 2014. Para o desenvolvimento do BrCris, o IBICT se inspirou num empreendimento de natureza semelhante em Portugal, o projeto PTCris, realizado pela Agência de Financiamento Portuguesa e pela Fundação para a Ciência e Tecnologia (FCT), também no contexto do euroCRIS (Bevan; Harrington, 2011; Jetten; Simons; Rijnders, 2019).

O BrCris é, portanto, um ecossistema que permite a interoperabilidade de dados científicos e tecnológicos provenientes de diferentes origens. Esse ecossistema integra dados de várias fontes, como agências de fomento, bancos de dados científicos e técnicos, repositórios institucionais, plataformas governamentais, infraestrutura de pesquisa, bancos de dados de teses e dissertações etc. (Figura 2). Dada a heterogeneidade dos sistemas a serem integrados à proposta do BrCris, o principal desafio é tornar possível a reunião de forma consistente dos dados e metadados da pesquisa de Ciência e Tecnologia (C&T) brasileira em uma única base.

2 Para mais informações, consultar: <https://eurocris.org/community/members>.

Figura 2 – O ecossistema do universo de desenvolvimento BrCris



Fonte: dados da pesquisa (2024).

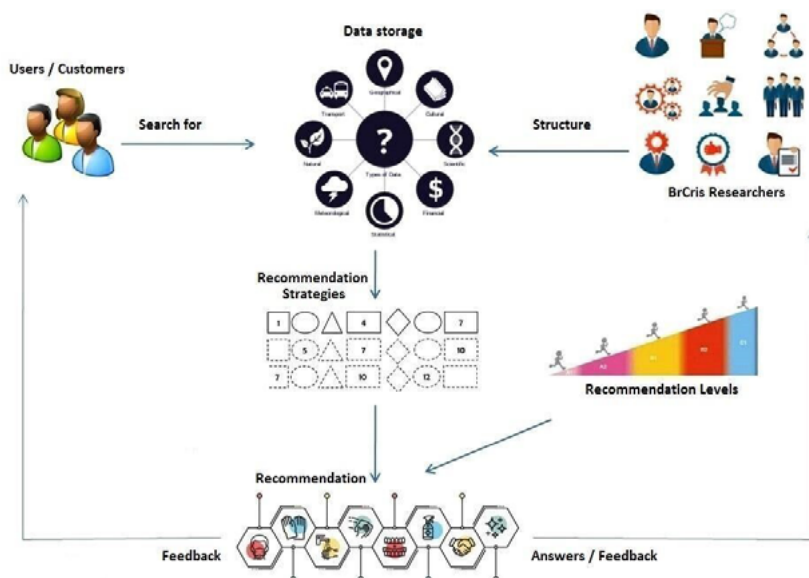
Vários sistemas oferecem funcionalidades similares às propostas pelo BrCris. Dentre eles, podemos citar agregadores de publicações científicas como Web of Science, Scopus e Google Scholar, redes de pesquisadores como ResearchGate e Microsoft Academia.edu, e bases de dados como Dimensions e OpenAIRE Research Graph, que oferecem redes mais sofisticadas envolvendo não só pesquisadores e publicações como também diversos outros elementos do domínio de pesquisa científica – *datasets*, financiamentos, patentes, ensaios clínicos etc. Dentro desse universo de opções, a proposta de criação do BrCris surge da necessidade de um sistema de informações científicas livre de vícios comerciais, característica de vários dos sistemas citados, e que esteja totalmente aberto à cooperação da comunidade acadêmica brasileira, não se configurando, portanto, como um produto fechado desenhado unilateralmente, mas uma solução colaborativa que visa ouvir e atender às necessidades dos diferentes atores da área de desenvolvimento científico e tecnológico no Brasil.

Desenvolvimento do BrCris

O principal objetivo da construção do BrCris é fornecer um sistema que permita, em um único ambiente, operar e monitorar a interoperabilidade dos dados de ciência e tecnologia brasileiros (Figura 3). Esse sistema é capaz de:

- analisar conjuntos de dados recuperados pelos sistemas englobados pelo projeto;
- tratar, com o apoio de representações, os conjuntos de dados recuperados para realizar a deduplicação, desambiguação, identificação de entidades, classificação, geolocalização, além das relações e da aplicação de métricas;
- disponibilizar um banco de dados padronizado e organizado por segmentos da ciência e tecnologia, tratados semanticamente e com certificação de dados;
- disponibilizar ferramentas para pesquisa e visualização dos quatro níveis de Sistemas de Recomendação por meio das métricas de classificação de aplicativos.

Figura 3 – Modelo representativo dos Sistemas de Recomendação oferecidos pelo BrCris



Fonte: adaptada de Schafer, Konstan e Riedl (2001).

O estudo para o desenvolvimento do BrCris e seus Sistemas de Recomendação foi realizado a partir de duas dimensões. A primeira dimensão reflete o que seria ideal em Sistemas de Recomendação, e a segunda reúne todo um conjunto de dados qualificados para servir como Sistemas de Recomendação. Com base em Schafer e demais autores (2001), foi considerado um sistema ideal para fixar informações estratégicas para qualquer usuário, mas com redução de escopo com base no que seria priorizado nos Sistemas de Recomendação do BrCris.

A segunda dimensão foi estruturada em três partes apoiadas em módulos desenvolvidos em Python e na plataforma LA Referencia para orquestração da carga, processamento e exportação de dados. A LA Referencia oferece um portal que visa dar visibilidade à produção científica de instituições de ensino superior e pesquisa na América Latina, promovendo o acesso livre e aberto ao texto completo, e provê também um conjunto de ferramentas para tratamento desses dados. Os processos estão organizados da seguinte forma:

- Mapeamento de um conjunto de plataformas e bancos de dados, incluindo Oasisbr, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), Plataforma Lattes, Espacenet, WikiData, Diadorim, Latindex, Directory of Open Access Journals (DOAJ), UlrichsWeb, ROR, Dimensions, Altmetric, Plataforma Sucupira, OpenAIRE, DBpedia, OpenCitations, Sherpa Romeo e Google Scholar. Nesse conjunto, as ferramentas utilizadas são para coleta, homogeneização e verificação de dados;
- Organização das entidades identificadas no conjunto de dados. Esse processo visa à identificar os projetos, serviços, programas de pós-graduação, cursos, organizações e pesquisadores, além de patentes, publicações e conjuntos de dados. Por meio desse processo, verificamos também como todas essas entidades estão relacionadas;
- Pesquisa, visualização, armazenamento, exportação e interoperabilidade de dados por meio do uso de APIs da plataforma LA Referencia. Esses processos permitem a elaboração do contexto semântico do universo de dados.

Os principais resultados são uma API para recuperação de dados por outras aplicações, uma interface de busca e navegação, e um conjunto de *dashboards* que permitem vislumbrar a organização e a representação dos dados científicos e tecnológicos brasileiros. Quanto à parte prática, utilizou-se uma abordagem mista, quantitativa e qualitativa, testando tecnologias e conjuntos de dados heterogêneos de diferentes naturezas, criando um mecanismo de operacionalidade, visualização e aplicação de métricas e adotando uma forte base experimental com testes contínuos. Com isso, visamos à criação de um sistema inovador e único no contexto de procedimentos de pesquisa no Brasil.

Modelos de representação

Os esquemas de representação de dados permitem a definição do escopo do projeto, por intermédio da delimitação dos tipos de entidades que pretendemos representar. O BrCris adota dois níveis de representação: *lógico* e

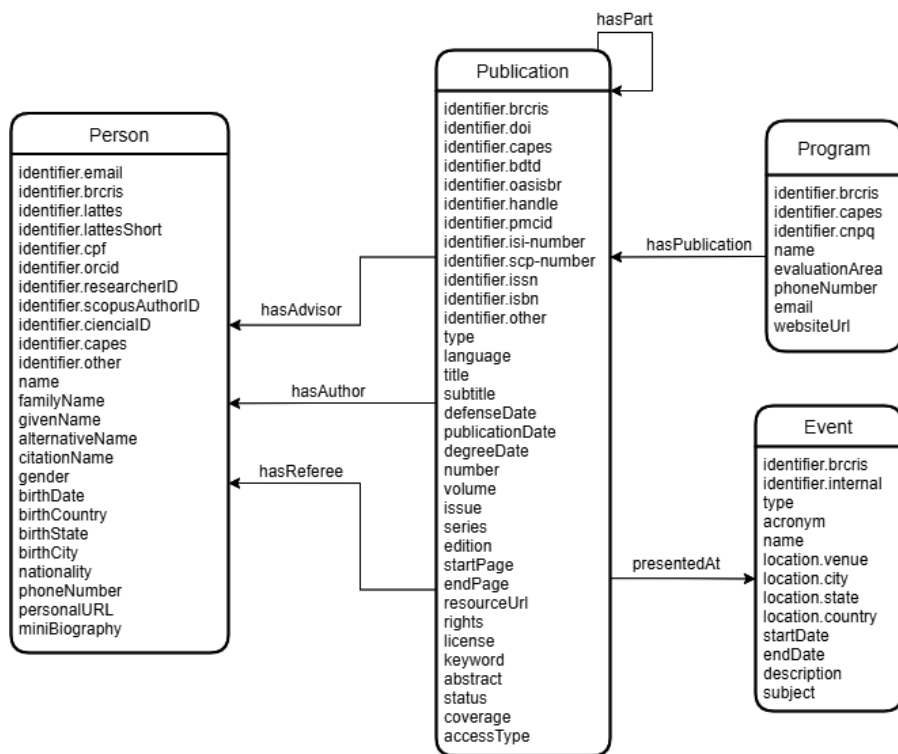
semântico. O nível lógico é materializado por via de um Modelo de Entidade e Relacionamento, também chamado de metamodelo, baseado no Common European Research Information Format (Cerif) (Jörg, 2009). Não se trata de um modelo relacional, mas de uma representação menos rígida composta por um conjunto de entidades contendo cada uma um conjunto de atributos, e ligadas por um conjunto de relações binárias. É um formato de representação interna elaborado para servir como base para o esquema físico do banco de dados na plataforma LA Referencia. Isso significa que o metamodelo é traduzido para um esquema de banco de dados físico, mas é totalmente configurável, sendo facilmente estendido à medida que novos tipos de dados são agregados ao projeto, sem afetar as estruturas internas da plataforma LA Referencia.

As entidades incluídas no metamodelo são: *Person, Publication, OrgUnit, Journal, Event, Service, Course, GraduateProgram, Patent, Project, Funding, ResearchGroupMembership, ProjectMembership, Authorship, Advisoring, Refereeing, Editing e ResearchArea*. A Figura 4 ilustra, de forma simplificada, parte desse modelo, com foco na entidade *Publication* e alguns de seus relacionamentos.

O Modelo de Entidade e Relacionamento é, portanto, responsável por organizar e integrar toda essa informação para que ela possa de fato ser utilizada para os objetivos do projeto. Entretanto, esse formato é destinado a atender às demandas de armazenamento interno dos dados, sendo necessário um formato mais padronizado para fins de acesso, interpretação e reuso dos dados por agentes externos.

O nível de representação semântico foi pensado como uma forma de disponibilizar os dados de forma padronizada, acessível e reutilizável, e foi implementado como uma ontologia que permite a visualização e a navegação dos dados como um grafo de conhecimento. A base desse modelo é a ontologia VIVO (Corson-Rikert *et al.*, 2012), um modelo que descreve o domínio de pesquisa acadêmica e é composto de classes e propriedades para representar uma rede de pesquisadores, as instituições e projetos aos quais estão associados, e as publicações e demais produtos resultantes de suas pesquisas, como patentes, *softwares*, *datasets* etc.

Figura 4 – Parte do Metamodelo de Entidade e Relacionamento



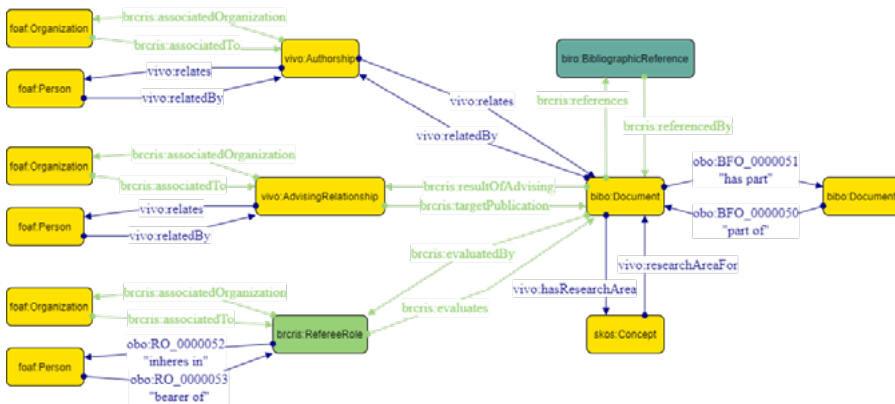
Fonte: dados da pesquisa (2024).

O ponto de partida para a criação do modelo semântico foi o Modelo de Entidade e Relacionamento. Cada um dos elementos desse modelo – entidades, atributos e relações – foi mapeado para um elemento da ontologia VIVO, mediante uma busca sistemática para identificação de elementos equivalentes. Quando uma classe ou uma propriedade equivalente não foi encontrada, um novo elemento foi criado, dando origem a uma extensão local da ontologia, identificada pelo *namespace* “*brcris*”. O modelo semântico final é, portanto, formado por um subconjunto da ontologia VIVO mais uma extensão local destinada a acomodar informações específicas do ambiente de pesquisa científico brasileiro.

A classes incluídas no modelo semântico, com seus respectivos *namespaces*, são: *foaf:Person*, *foaf:Organization*, *obo:ERO_0000005 (Service)*, *bibo:Journal*,

bibo:Document, *vivo:Course*, *bibo:Patent*, *vivo:Project*, *vivo:Grant*, *event:Event*, *biro:BibliographicReference*, *vivo:MemberRole*, *vivo:Authorship*, *vivo:AdvisingRelationship*, *vivo:Editorship*, *skos:Concept*, *brcris:Community*, *brcris:GraduateProgram* e *brcris:RefereeRole*. A Figura 5 mostra uma parte do modelo final, centrada na classe *bibo:Document* – que corresponde à entidade *Publication* – e alguns de seus principais relacionamentos. As setas azuis representam propriedades de objetos provenientes da ontologia VIVO, que têm sempre uma propriedade inversa. Esse mesmo padrão foi seguido na criação de novas propriedades na extensão local da ontologia (setas verdes).

Figura 5 – Parte do modelo semântico



Fonte: dados da pesquisa (2024).

Coleta e tratamento dos dados

De acordo com os objetivos do projeto BrCris, as fontes de informação coletadas concentram-se principalmente em: (a) teses e dissertações que foram produzidas por pesquisadores brasileiros; (b) produtos de ciência aberta fornecidos por repositórios de dados e bancos de dados acadêmicos/científicos; (c) bases de dados de patentes; (d) sistemas métricos científicos; (e) plataformas de perfis de pesquisa; (f) sistemas de avaliação científica; e (g) diretórios de revistas científicas.

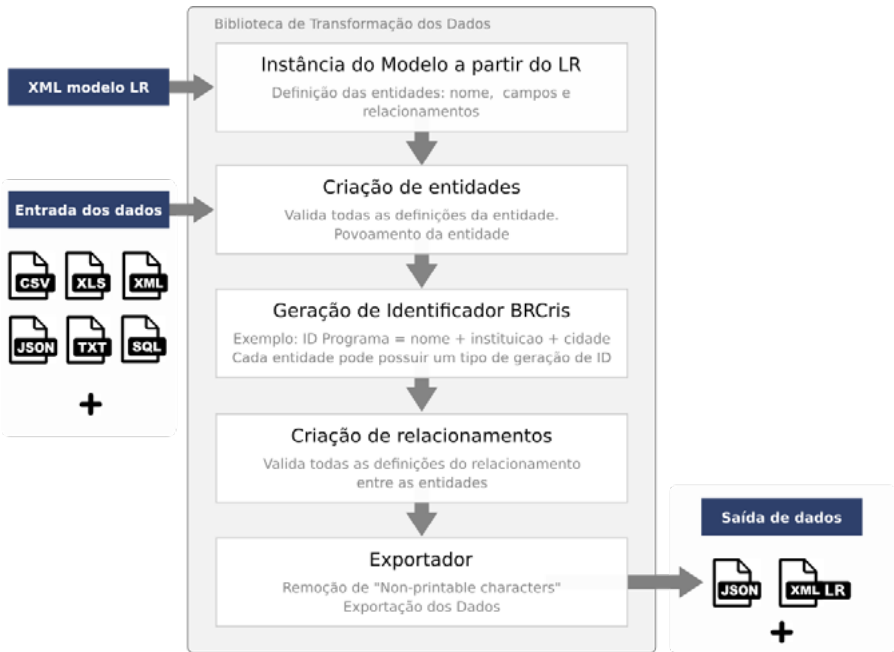
O BrCris reúne, portanto, dados de diversas origens, sendo que, inicialmente, foram selecionados os principais repositórios de dados que seriam

objeto de interesse para o projeto. As fontes de dados coletadas e as entidades extraídas de cada uma delas são:

- Plataforma Lattes: Dados curriculares, sendo considerada a principal fonte de dados do BrCris, trazendo dados pessoais dos indivíduos, local de trabalho, publicações, eventos, cursos e revistas. Esses dados foram transformados nas seguintes entidades do modelo: *Person*, *OrgUnit*, *Publication*, *Event*, *Program*, *Course* e *Service*;
- Oasisbr: Repositório mantido pelo IBICT que fornece dados confiáveis sobre publicações científicas em acesso aberto. Os dados foram mapeados para as entidades *Publication*, *Journal*, *Service* e *Person*;
- BDTD: Assim como o Oasisbr, a BDTD também é mantida pelo IBICT. Fornece dados confiáveis sobre teses e dissertações brasileiras. Os dados foram mapeados para as entidades *OrgUnit*, *Publication*, *Course* e *Person*;
- Plataforma de Instituições em Ciência, Tecnologia e Inovação (PCTI): Também é desenvolvida pelo IBICT, no contexto do projeto BrCris. Os dados foram mapeados para a entidade *OrgUnit*;
- Plataforma Sucupira: Dados de programas de pós-graduação que fornecem uma base confiável de programas brasileiros. Os dados foram mapeados para as entidades *Program*, *Publication*, *Course* e *OrgUnit*;
- Instituições do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep): Assim como os programas da Plataforma Sucupira, o Inep fornece uma base confiável, mas de instituições, sendo mapeados para a entidade *OrgUnit*.
- API de Dados Abertos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes): Dados confiáveis fornecidos pela Capes, mapeados para diversas entidades, como *Person*, *OrgUnit*, *Publication*, *Program* e *Course*.
- Revistas: O processamento de dados sobre revistas científicas fornece um conjunto também confiável, sendo mapeados para a entidade *Journal*. Esses dados são provenientes das plataformas Diadorim, Latindex, DOAJ e UlrichsWeb.

Para a parte do tratamento dos dados recuperados, foi criada uma representação para obtenção, transformação e processamento de dados. Um extrator incremental que trabalha com uma listagem de identificadores utilizando o protocolo Simple Object Access Protocol (SOAP) foi utilizado para acessar e extrair os currículos da Plataforma Lattes em formato XML. Também foi criado um documento descritivo com todas as entidades e relacionamentos possíveis dentro dos conjuntos de dados recuperados das plataformas e dos bancos de dados. Outra criação arquitetônica foi a estruturação dos dados no formato XML esperado pela plataforma LA Referencia. Para isso, uma biblioteca foi desenvolvida em Python para identificar nos dados de origem os elementos de interesse e transformá-los nas entidades, nos atributos e nos relacionamentos a serem armazenados. Outra função dessa biblioteca é a geração de identificadores únicos do BrCris para cada uma das entidades extraídas (Figura 6).

Figura 6 – Esquema de coleta e tratamento de dados no BrCris



Fonte: dados da pesquisa (2024).

Para a carga e o processamento dos dados, é utilizada a plataforma LA Referencia. Essa plataforma é implementada na linguagem de programação Java, tem mais de sete anos de desenvolvimento, e mais de dez países a utilizam com o mesmo propósito. A LA Referencia trabalha com coleta de dados de repositórios por meio do protocolo de coleta de metadados OAI-PMH e atende às demandas de validação e transformação de metadados do BrCris.

No contexto do BrCris, a plataforma LA Referencia é responsável por agregar os dados provenientes das diversas fontes de informação. Ela armazena, processa e identifica entidades equivalentes fazendo o uso de procedimentos de deduplicação, além de permitir a recuperação e a exportação dos dados para diversos formatos.

Portanto, o repositório de dados do BrCris é composto por diversas fontes integradas e padronizadas. Esse repositório pode ser usado para consultas e análises usando ferramentas de visualização. Os dados também podem ser exportados usando APIs disponíveis em um formato padronizado e certificado de conteúdo. Por exemplo, um conjunto de dados que poderia ser usado tanto para análise bibliométrica quanto consumido por outras aplicações seria o conjunto de publicações de um determinado grupo de indivíduos. Na Figura 7, observamos o conteúdo de um arquivo gerado pela biblioteca Python do BrCris. Esse arquivo contém publicações de um pesquisador específico.

Figura 7 – Fragmento de dado com as publicações de um pesquisador

```

<entity type="Publication" ref="Publication46">
  <semanticIdentifier>bcrcis:f9c64c8d702f238918ff8b9a6665e</semanticIdentifier>
  <field name="title" value="On interdisciplinary collaborations in scientific coauthorship networks: the case of the Brazilian community"/>
  <field name="language" value="Inglês"/>
  <field name="publicationDate" value="2020"/>
  <field name="volume" value="124"/>
  <field name="series" value="003"/>
  <field name="startPage" value="2341"/>
  <field name="endPage" value="2360"/>
  <field name="type" value="article"/>
  <field name="keyword" value="Bibliometria"/>
  <field name="keyword" value="Cientometria"/>
  <field name="keyword" value="Colaboração Científica"/>
  <field name="keyword" value="Plataforma Lattes"/>
  <field name="resourceUrl" value="http://https://link.springer.com/article/10.1007/s11192-020-03605-3"/>
  <field name="identifier.brcris" value="f9c64c8d702f238918ff8b9a6665e"/>
  <field name="author" value="GERALDO J PESSOA JUNIOR"/>
  <field name="author" value="THIAGO MAGELA RODRIGUES DIAS"/>
  <field name="author" value="THIAGO H P SILVA"/>
  <field name="author" value="ALBERTO HENRIQUE FRADE LAENDER"/>
</entity>
<entity type="Journal" ref="Journal17">
  <semanticIdentifier>bcrcis:439df380a32caf86c6ef8868d95a3192</semanticIdentifier>
  <semanticIdentifier>issn:1588-2861</semanticIdentifier>
  <field name="title" value="SCIENTOMETRICS (ONLINE)"/>
  <field name="type" value="journal"/>
  <field name="identifier.issn" value="1588-2861"/>
  <field name="identifier.brcris" value="439df380a32caf86c6ef8868d95a3192"/>
</entity>
<entity type="Person" ref="Person11">
  <semanticIdentifier>lattes:9089204821424223</semanticIdentifier>
  <field name="name" value="ALBERTO HENRIQUE FRADE LAENDER"/>
  <field name="citationName" value="LAENDER, A. H. F."/>
  <field name="identifier.lattes" value="9089204821424223"/>

```

Fonte: dados da pesquisa (2024).

Como pode ser observado, todas as publicações relacionadas a um determinado pesquisador são mapeadas para o modelo proposto. Todos os dados são corretamente classificados e padronizados, possibilitando sua importação por outras ferramentas de análise ou repositórios locais. Os registros agregam informações sobre a própria publicação – entidade do tipo *Publication* – incluindo a geração de um identificador semântico, as informações sobre o periódico em que o trabalho foi publicado – entidade do tipo *Journal* – e sobre o conjunto de autores que participaram da publicação – entidades do tipo *Person*. Ressalta-se que todos os identificadores persistentes atribuídos a cada uma das entidades são incluídos nos dados da entidade, possibilitando a desambiguação dos dados, o que é considerado um grande desafio na integração de grandes repositórios em diferentes formatos³.

3 Arquivos contendo dados de publicações no modelo adotado pelo BrCris podem ser encontrados em <https://doi.org/10.5281/zenodo.4955756>.

Disponibilização dos dados

A plataforma LA Referencia, principal ferramenta de apoio no processo de tratamento das informações do BrCris, oferece um conjunto de funcionalidades para o processamento de dados oriundos de diversas origens, permitindo o tratamento de conjuntos de dados heterogêneos para a criação de uma base deduplicada e unificada. Após estes serem carregados, deduplicados e consolidados, eles podem ser exportados para outras ferramentas para consulta, visualização e análise.

Os procedimentos de exportação são implementados como rotinas de Extração, Transformação e Carga (ETL), permitindo que os dados relacionais sejam convertidos para diferentes formatos. O módulo de Extração é compartilhado por todos os exportadores/indexadores, e é responsável por recuperar na base relacional as entidades por tipo – por exemplo, *Person*, *Publication*, *Patent* etc. Para cada entidade, são recuperados também seus conjuntos de atributos e relacionamentos com outras entidades. É possível extrair todos os tipos de entidade de uma única vez, ou especificar um único tipo a ser extraído isoladamente.

Os módulos de Transformação e Carga são específicos para cada exportador/indexador. A transformação é baseada em arquivos de configuração que descrevem o mapeamento entre os elementos de origem – entidades, atributos e relacionamentos – e os objetos de destino, e conta, assim como a carga, com o apoio de APIs específicas para as ferramentas e formatos de destino (JSON, RDF etc.). Os principais exportadores/indexadores disponibilizados pela plataforma LA Referencia atualmente são para o Solr, o Elasticsearch e o VIVO.

O indexador para o Solr permite a criação de um índice por tipo de entidade ou um único índice contendo diferentes tipos de entidades. Na configuração de transformação, cada atributo de uma entidade é mapeado para um campo no índice, quanto aos atributos de entidades relacionadas, estes também podem ser definidos como campos. Assim, na indexação de uma entidade do tipo *Publication*, por exemplo, é possível definir como campos do índice, além de quaisquer atributos da própria publicação, o nome de seus autores – atributo das entidades do tipo *Person* relacionadas a essa

publicação como autores –, o Número Internacional Normalizado para Publicações Seriadas (ISSN) do periódico onde foi publicada – atributo da entidade do tipo *Journal* – etc.

Os dados indexados no Solr são então disponibilizados para consulta por meio de uma API REST, podendo ser consumidos por outras aplicações. A Figura 8 mostra, na interface Swagger, o *endpoint* para consulta de publicações. Informando-se o ID Lattes ou o ID de pesquisador e colaborador aberto (Orcid) de uma pessoa – além de outros parâmetros opcionais –, são retornadas, em formato JSON, as informações de todas as publicações às quais esta pessoa está relacionada, seja por autoria, orientação ou coorientação – no caso de teses e dissertações. Atualmente, a principal aplicação da API é na certificação de teses e dissertações declaradas por pesquisadores na Plataforma Lattes, conforme descrito na próxima seção.

Figura 8 – *Endpoint* para consulta de publicações na API do BrCris

GET

/search/results/publications

Returns a list of entities of type Publication

Parameters

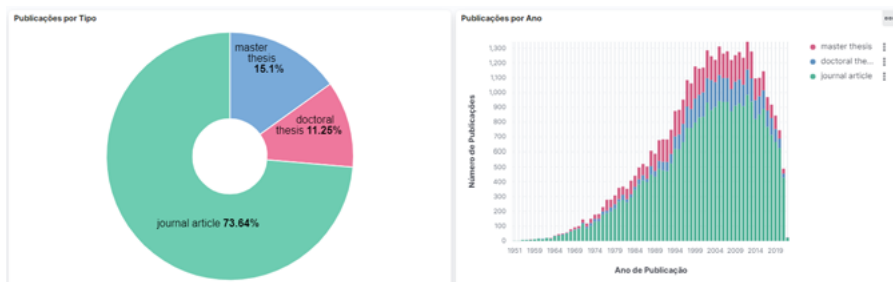
Name	Description
communityId string (query)	BrCris internal Community ID communityId - BrCris internal Community ID
page (query)	Results page you want to retrieve (0..N) page - Results page you want to retrieve (0..
personId string (query)	Author, advisor, or co-advisor Lattes ID or Orcid personId - Author, advisor, or co-advisor Latti
publicationDate string (query)	Publication Date - Use YYYY publicationDate - Publication Date - Use YYYY
size (query)	Number of records per page. size - Number of records per page.
title string (query)	Publication title - Use normalized title: lower case, no stopwords, no accents, no punctuation title - Publication title - Use normalized title: li
type string (query)	Publication type, e.g master thesis, doctoral thesis, conference proceedings, etc. type - Publication type, e.g master thesis, do

Fonte: dados da pesquisa (2024).

O indexador do Elasticsearch funciona de forma semelhante ao do Solr, utilizando um arquivo de configuração que mapeia os elementos de origem para os campos de um ou mais índices de destino. A principal diferença em termos de formato é que, enquanto os registros no índice do Solr são *flat* – ou seja, uma lista de campos em um único nível –, os índices do Elasticsearch permitem entidades aninhadas, gerando, portanto, uma estrutura hierárquica de campos. Dessa forma, uma entidade do tipo *Publication* indexada no Elasticsearch pode ter um campo chamado *authors*, cujo valor é uma lista de entidades do tipo *Person*, cada qual com seu próprio conjunto de campos.

Os dados indexados no Elasticsearch são consumidos pela ferramenta de visualização Kibana, que permite a criação de *dashboards* que agregam diferentes tipos de relatórios – tabelas, gráficos etc. – para análise quantitativa dos dados. A Figura 9 mostra exemplos de visualizações em um *dashboard* criado no Kibana a partir dos dados de publicações de pesquisadores sênior do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Figura 9 – Gráficos no *dashboard* de publicações de pesquisadores sênior no Kibana



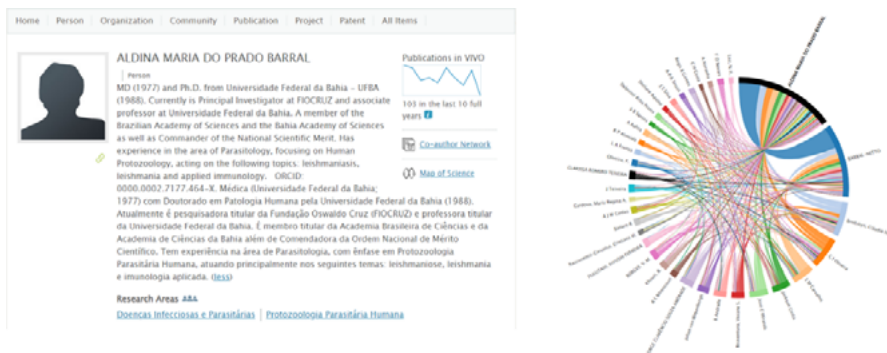
Fonte: dados da pesquisa (2024).

Por fim, o exportador do VIVO tem como finalidade gerar um grafo em RDF para navegação e exploração dos dados nessa ferramenta. A transformação é baseada em um mapeamento entre as entidades, os atributos e os relacionamentos recuperados e os elementos do modelo semântico descrito anteriormente, tratando-se, portanto, de um processo de população de ontologia. Cada elemento retornado pelo módulo de Extração é mapeado para uma ou mais triplas RDF. Essas triplas são adicionadas a um modelo

RDF que é enviado para carga na base do VIVO por intermédio de sua API SPARQL (Börner *et al.*, 2012).

Diferentemente dos indexadores do Solr e do ElasticSearch, nos quais normalmente apenas um tipo de entidade é indexado por vez, no VIVO todos os tipos de entidades são exportados de uma única vez, porque o objetivo dessa interface é permitir a navegação entre todos os tipos de objetos que, por estarem organizados em um grafo de conhecimento, oferecem uma visão geral do universo de informações tratadas. A Figura 10 mostra um perfil de usuário e sua rede de coautores, algumas das visualizações possíveis na plataforma VIVO.

Figura 10 – Perfil de usuário e rede de coautores no VIVO



Fonte: dados da pesquisa (2024).

Ao disponibilizar os dados em diferentes ferramentas e formatos, o BrCris permite que as informações do ecossistema de pesquisa brasileiro sejam exploradas em diferentes dimensões, atendendo a finalidades distintas e complementares.

Desdobramentos

Além do projeto BrCris em si, o IBICT vem implementando recentemente outros projetos de desenvolvimento de infraestruturas informacionais, que servirão também como base para o pleno desenvolvimento do BrCris. Os principais são o Projeto Laguna - Infraestrutura Informacional Aberta

e Lago de Dados, e o Projeto CEOIS - Plataforma Inteligente de Coleta, Organização, Certificação, Análise e Reuso da Informação. Além destes, outros projetos resultantes de parcerias com diferentes instituições utilizam o BrCris como fonte de informação ou como referência para a criação de funcionalidades em sistemas existentes, ou de novos sistemas especializados para agregação de informações de pesquisa científica.

Outras iniciativas

Várias iniciativas recentes podem ser consideradas como desdobramentos do projeto BrCris. Uma delas é a parceria entre o IBICT e a Fundação Oswaldo Cruz (Fiocruz) para aprimoramento do Observatório em Ciência, Tecnologia e Inovação em Saúde⁴. Essa cooperação tem como finalidade a aplicação das técnicas e da reutilização dos dados tratados no contexto do BrCris para aumentar a robustez na geração dos indicadores em Ciência, Tecnologia e Inovação (CT&I) disponibilizados pelo Observatório da Fiocruz.

A certificação de teses e dissertações no Currículo Lattes é outro importante desdobramento. Consumindo, por via da API REST BrCris – descrita anteriormente –, dados resultantes do cruzamento entre as informações de currículos e do Oasisbr, a Plataforma Lattes pode verificar se uma tese ou dissertação declarada por um pesquisador tem um registro correspondente em um repositório institucional oficial, podendo, dessa forma, conferir à publicação um selo de certificação. Também em parceria com a Plataforma Lattes, está previsto o autopopulamento do LattesData, mediante as informações de conjuntos de dados de pesquisa a serem coletados e tratados pelo BrCris.

O DFCRIS, também desenvolvido em parceria com o IBICT, é um derivado do BrCris, com o intuito de mapear as informações do ecossistema de pesquisa científica das instituições do Distrito Federal. Esta é uma colaboração de mão dupla, já que o DFCRIS, além de aproveitar a experiência já adquirida no desenvolvimento do BrCris, por ter um escopo mais reduzido, permite a experimentação de novos métodos e tecnologias, que serão posteriormente aplicados de volta no contexto nacional.

4 Para mais informações, consultar: <https://observatorio.fiocruz.br/>.

Futuros desenvolvimentos e parcerias incluem o Observatório de Ciência Aberta (OCABr), projeto coordenado pelo CNPq para a implantação de observatório para o progresso de práticas de ciência aberta no Brasil e, a Torre MCTI, iniciativa do Ministério da Ciência, Tecnologia e Inovações também voltada à representação de redes de instituições e especialistas, plataformas digitais, banco de dados e outros sistemas vinculados à CT&I.

O BrCris como base para Sistemas de Recomendação

As possibilidades de geração de Sistemas de Recomendação são ilimitadas. Entretanto, no escopo do BrCris, apresentamos quatro modelos com dados complementares e ações independentes. A primeira possibilidade é um sistema para identificar especialistas em um determinado assunto com base em sua produção científica. As fontes de dados utilizadas para análise de sua produção são Diadorim, Latindex, Portal ISSN, DOAJ, UlrichsWeb, Plataforma Lattes e Google Scholar Metrics. Essa análise, baseada em métricas, é focada em:

- tempo de resposta desde o envio até a publicação;
- média h-5;
- mediana h-5;
- taxa de publicação;
- comitê de revistas – nacional ou internacional;
- endogenia para periódicos brasileiros;
- idiomas de publicação.

Os resultados dessa análise podem ser úteis para as seguintes finalidades:

- diferenciar a evolução das áreas de conhecimento;
- verificar o relacionamento entre as áreas de conhecimento;
- identificar a junção de terminologias em diferentes áreas;
- identificar a relevância informacional;

- visualizar as tendências de crescimento e queda das publicações;
- agrupar áreas de pesquisa ou pesquisadores a partir da análise de resultados e citações;
- verificar como é formado cada universo de dados e como dividir as grandes áreas do conhecimento;
- identificar áreas centrais e periféricas de pesquisa e sua evolução;
- visualizar a evolução das tipologias de publicação – artigos de periódicos, livros e capítulos de livros, textos em jornais e revistas, anais e apresentações etc.

Os tipos de visualização planejados para apoiar esse sistema de recomendação estão ilustrados na Figura 11.

Figura 11 – Visualizações do sistema de recomendação de produção científica



Fonte: The Data Visualisation Catalogue ([202-]).

A segunda ação é um sistema de recomendações para teses e dissertações. Sob a orientação da Plataforma Lattes, a BDTD é utilizada como fonte de dados. O foco dessa análise é identificar:

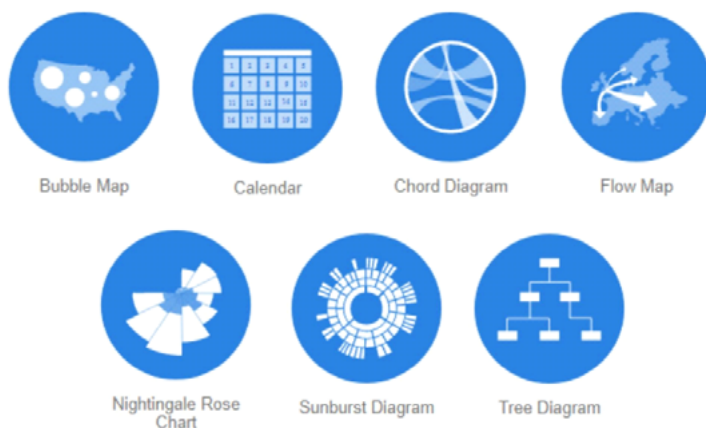
- genealogia em até 11 níveis;
- orientações por região;

- especialistas em áreas temáticas de orientação acadêmica;
- especialistas em áreas temáticas de participação em bancas.

As visualizações planejadas para esse sistema de recomendação (Figura 12) visam auxiliar as seguintes análises:

- averiguar a genealogia científica, visando saber quem são os pais em termos científicos no processo de formação de pesquisadores;
- identificar o acompanhamento científico, buscando saber se determinado pesquisador orientou um grupo de outros pesquisadores em todos os níveis de formação – graduação, mestrado e doutorado – ou em parte desses níveis;
- visualizar a origem das defesas em nível de regionalização, visando identificar defesas em outros países;
- identificar defesas fora do Brasil para verificar o andamento dessas formações e saber quais foram os principais destinos;
- identificar a frequência de defesas por período;
- analisar temporalmente os pesquisadores que mais orientaram por áreas e por temáticas.

Figura 12 – Visualizações do sistema de recomendação de teses e dissertações



Fonte: The Data Visualisation Catalogue ([202-]).

A terceira ação diz respeito a um sistema de recomendação para patentes e inovação, não podendo ser estendido à outra tipologia, pois a coleta de dados dessa natureza é complexa. As fontes são o banco de dados de patentes Espacenet e a Plataforma Lattes. Esse sistema visa analisar:

- citações recebidas;
- citações feitas;
- concessão e extensão;
- famílias de patentes;
- patentes triádicas;
- classificação de patentes.

As visualizações para esse sistema de recomendação (Figura 13) nos permitem identificar e analisar:

- citações feitas e recebidas;
- patentes protegidas em outros países e pedidos de registro de famílias de patentes – para determinar a triádica mundial;
- concessão e renovação de patentes – por via de um dendrograma de classificação e relacionamento;
- listas de inventores e depositantes, incluindo relação depositante-inventor;
- valores de patentes pela aplicação do valor econômico direto – tamanho da família –, valor econômico indireto – anuidade – e valor tecnológico – citações recebidas.

Figura 13 – Visualizações do sistema de recomendação para patentes



Fonte: The Data Visualisation Catalogue ([202-]).

Por fim, a quarta ação trata de um sistema de recomendação para editores acadêmicos que trabalham com bancos de dados e diretórios como Latindex, Portal ISSN, DOAJ, UlrichsWeb, Plataforma Lattes (*productivity Sscholars, doctors/masters*) e Google Scholar Metrics. Esse sistema se concentra em dados métricos, incluindo:

- média h-5;
- mediana h-5;
- produção acadêmica;
- produção mestre.

A representação de termômetro (Figura 14) é utilizada na visualização desses dados, indicando em que nível os editores considerariam submeter a revista aos principais canais de publicação.

Figura 14 – Visualizações do sistema de recomendação para editores



Fonte: The Data Visualisation Catalogue ([202-]).

Considerações finais

O BrCris é um projeto abrangente que tem como finalidade disponibilizar uma visão unificada do ecossistema de pesquisa científica brasileiro. Com o objetivo de reunir, integrar e inter-relacionar dados de diversas fontes distintas, um conjunto de técnicas e ferramentas tem sido testado e adotado na construção do sistema. Esse sistema abrange diferentes tipos de interfaces, voltadas para busca e navegação de dados, análises quantitativas por via de relatórios e *dashboards* e recomendação em diversas áreas dentro do domínio de pesquisa acadêmica.

Até o momento foram feitos avanços significativos em relação aos esforços de desenvolvimento. Podem ser contabilizados como produtos concretos desses esforços a elaboração dos modelos de representação dos dados, a implementação de técnicas de coleta e tratamento de conjuntos de dados heterogêneos – aperfeiçoadas por meio de diversas iterações de carga e deduplicação – e a disponibilização de interfaces para usuários, seja por um conjunto piloto de dados, como no VIVO e no Kibana, ou como uma ferramenta já disponível e utilizada em produção, como a API REST.

Os próximos passos envolvem a consolidação do sistema com o auxílio da coleta integral, e subsequente atualização incremental dos conjuntos de dados de todas as fontes selecionadas. Essa consolidação apresenta diversos desafios inerentes à manipulação de grandes volumes de dados heterogêneos, seja no processamento – em especial na deduplicação de entidades –, ou nos processos de transformação e carga que disponibilizam os dados para

visualização. Os novos projetos e parcerias que surgiram como desdobramentos do BrCris – principalmente aqueles com foco em Big Data como o Laguna – auxiliam na superação desses desafios à medida que possibilitam a continuidade da busca por melhores soluções e práticas.

Por fim, a concretização dos Sistemas de Recomendação com base nos dados consolidados no escopo do BrCris – também planejada como próximo passo no desenvolvimento do projeto – representa o acréscimo de uma nova dimensão às funcionalidades do sistema: além de visualização, descoberta e análise dos dados, o BrCris dará também suporte à tomada de decisões, sendo uma fonte confiável de métricas e indicadores para a comunidade acadêmica brasileira.

Referências

BARCELLOS, C. D. *et al.* Sistema de recomendação acadêmico para apoio à aprendizagem. *RENOTE*, Porto Alegre, v. 5, n. 2, dez. 2007. Disponível em: <https://seer.ufrgs.br/index.php/renote/article/view/14236/8152>. Acesso em: 7 out. 2024.

BEVAN, S.; HARRINGTON, J. Managing research publications: lessons learned from the implementation of a Current Research Information System. *Serials*, [s. l.], v. 24, n. 1, p. 26-30, Mar. 2011. Disponível em: <https://serials.uksg.org/articles/1077/files/submission/proof/1077-1-1186-1-10-20150224.pdf>. Acesso em: 7 out. 2024.

BÖRNER, K. *et al.* VIVO: A semantic approach to scholarly networking and discovery. *Synthesis lectures on the Semantic Web: Theory and Technology*, [s. l.], v. 7, n. 1, p. 1-178, 2012.

CAZELLA, S. C. *Aplicando a relevância da opinião de usuários em sistema de recomendação para pesquisadores*. 2006. (Tese de Doutorado) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/8424/000575704.pdf?sequence=1&isAllowed=>. Acesso em: 7 out. 2024.

CAZELLA, S. C. *et al.* Recomendação de objetos de aprendizagem empregando filtragem colaborativa e competências. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 20., 2009,

- Florianópolis. *Anais [...]*. Florianópolis: [s. n.], 2009. Disponível em: http://www.niee.ufrgs.br/eventos/SBIE/2009/conteudo/artigos/completos/61791_1.pdf. Acesso em: 18 ago. 2020.
- CORSON-RIKERT, J. *et al.* The VIVO ontology. In: BÖRNER, K. *et al.* *VIVO*. [S. l.]: Springer, 2012. p. 15-33.
- THE DATA Visualisation Catalogue. [S. l., 202-]. Disponível em: <https://datavizcatalogue.com/>. Acesso em: 7 out. 2024.
- GRAVA, A. P. *Sistema de recomendação de artigos científicos utilizando dados sociais*. 2016. Dissertação (Mestrado em Sistemas de Informação) – Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, 2016. Disponível em: https://www.teses.usp.br/teses/disponiveis/100/100131/tde-26072016-160726/publico/dissertacao_arthur_grava-versao_corrigida_ppgsi_v2.pdf. Acesso em: 7 out. 2024.
- JETTEN, M.; SIMONS, E.; RIJNDERS, J. The role of CRIS's in the research life cycle. A case study on implementing a FAIR RDM policy at Radboud University, the Netherlands. *Procedia Computer Science*, [s. l.], v. 146, p. 156-165, 2019. Disponível em: <https://www.sciencedirect.com/science/article/pii/S187705091930095X?via%3Dihub>. Acesso em: 7 out. 2024.
- JÖRG, B. The common european research information format model (CERIF). CRISs for the European e-Infrastructure. *Data Science Journal*. [s. l.], 2009.
- KREBS, L. M.; ROCHA, R. P. da; RIBEIRO, C. Quem leu este também leu...: sistema de recomendação na biblioteca universitária. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 22, n. 1, p. 151-169, jan./mar. 2017. Disponível em: <https://www.scielo.br/j/pci/a/SDGHfzxHppxDDPcm6GmW8SD/?format=pdf&lang=pt>. Acesso em: 7 out. 2024.
- PINTO, A. L. *et al.* Brazil De-veloping Current Research Information Systems (BrCRIS) as data sources for studies of research. *Iberoamerican Journal of Science Measurement and Communication*, [s. l.], v. 2 n. 1, p. 1-12, 2022. DOI 10.47909/ijsmc.135. Disponível em: <https://ijsmc.pro-metrics.org/index.php/i/article/view/9/10>. Acesso em: 7 out. 2024.

RESNICK, P.; VARIAN, H. R. (ed.). Recommender systems. *Communications of the ACM*, New York, v. 40, n. 3, p. 56-58, Mar. 1997. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/245108.245121>. Acesso em: 7 out. 2024.

SCHAFER, J. B.; KONSTAN, J. A.; RIEDL, J. E-commerce recommendation applications. *Data mining and knowledge discovery*, [s. l.], v. 5, n. 1, p. 115-153, 2001. Disponível em: <https://doi.org/10.1023/A:1009804230409>. Acesso em: 7 out, 2024.

SCHAFER, J. B.; KONSTAN, J.; RIEDL, J. Recommender systems in e-commerce. *Proceedings of the 1st ACM conference on Electronic commerce 1999*, [s. l.], p. 158-166, [20--]. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/336992.337035>. Acesso em: 7 out. 2024.

SILVA, H. G. A. da. *Um estudo de caso de aplicação de sistemas de recomendação baseado em filtragem colaborativa na tarefa de designação de profissionais para serviços de TIC Offshore*. 2020. Dissertação (Mestrado em Engenharia de Produção e Sistema Computacionais) – Instituto de Ciência e Tecnologia, Universidade Federal Fluminense, Rio das Ostras, 2020. Disponível em: <https://app.uff.br/riuff/bitstream/handle/1/21694/Disserta%c3%a7%c3%a3o%20catalogada%20hugo%20guilherme.pdf?sequence=1&isAllowed=y>. Acesso em: 7 out. 2024.

CAPÍTULO 5

Google Scholar como infraestrutura de avaliação científica: potencialidades e limites¹

Fabiana Andrade Pereira

Rogério Mugnaini

Introdução

No conjunto das atividades de investigação, a produção científica se destaca como o mais importante elemento pelo qual a comunidade científica demonstra a pertinência e a relevância dos resultados de pesquisa. Diversos estudos bibliométricos têm seus dados mensurados em fontes de informação tradicionais, como as bases de dados. Para serem indexados nessas bases, os periódicos científicos passam por processos seletivos rigorosos, fato que contribui para a limitação de cobertura em tais bases. Além disso, essas disparidades também são ocasionadas ao mensurar o impacto científico em algumas áreas, devido às diferentes dinâmicas da produção científica.

Nota-se que a internet possibilita o acesso a outros tipos de comunicação científica, tidos como canais informais para divulgação dos resultados da pesquisa que, ainda em etapas preliminares, são passíveis de comunicação. Destaca-se, dessa forma, que os avanços tecnológicos, além de dissipar a percepção dos limites entre a comunicação científica formal e informal, também foram responsáveis por atender às novas necessidades, evidenciando as limitações das bases de dados tradicionais, como destaca Mugnaini (2018): “a ausência de parte significativa da literatura local em fontes de informação de abrangência mundial” fica patente a partir do momento em que outras

1 O capítulo é resultado da dissertação de mestrado de Fabiana Pereira, orientada por Rogério, defendida em 2022.

bases (locais) são analisadas em conjunto. Assim, justifica-se a criação e o uso de outras bases de dados para se empreender análises de desempenho.

A principal razão para o desenvolvimento da área de bibliometria se deve à disponibilidade de extensas fontes bibliográficas e bases de dados. Concomitante à diversidade infraestrutural, a bibliometria se tornou uma ferramenta padrão de política científica e gestão de pesquisa nas últimas décadas (Glänzel, 2003). Por isso, observa-se a relação próxima entre o desenvolvimento da infraestrutura e das abordagens avaliativas voltadas para a medição de desempenho (Krüger; Petersohn, 2022).

Apoiando essa ideia, Torres-Salinas e Jiménez-Contreras (2012) dizem que as competências básicas das práticas cientométricas e bibliométricas estão muito relacionadas ao conhecimento especializado de fontes de informação. Portanto, o controle dessas fontes de informação é um dos três pilares fundamentais para o desenvolvimento da área, pois a boa operação e o bom desempenho dos processos *downstream* dependem também das boas medidas das informações.

Verifica-se, portanto, a importância das fontes de informação, pois:

- a qualidade da informação é um dos componentes mais importantes para as práticas de avaliação dos impactos científicos. Portanto, quaisquer fontes de dados e informações que viabilizem a mensuração de resultados necessitam que seus filtros e parâmetros estejam firmados em critérios consistentes;
- é por meio dessas infraestruturas, aliadas à alta capacidade computacional para o armazenamento e acesso rápido aos dados bibliográficos, que surgem elementos capazes de medir a produção científica, viabilizando a disponibilidade de indicadores e ferramentas de análise.

Para Krüger e Petersohn (2022), no processo de medição de desempenho, as infraestruturas não devem ser entendidas como um suporte técnico neutro, visto que as ferramentas, funcionalidades, configurações, classificações e demais particularidades podem influenciar nos resultados avaliativos. Essas infraestruturas, seus provedores e seus usuários influenciam em como

a ciência e a pesquisa são compreendidas e como o desempenho da pesquisa é medido, atenta Krüger (2020).

Com o surgimento de infraestruturas digitais para publicação, indexação e gerenciamento de publicações científicas, desencadeou-se um crescimento sem precedentes na produção de dados bibliométricos, livremente acessíveis por bancos de dados. Os provedores de infraestruturas bibliométricas, como o Clarivate Analytics, o Elsevier ou o Digital Science, por exemplo, procuram constantemente coletar maiores quantidades de dados e metadados sobre publicações com melhor qualidade de dados para propor novas formas de usá-los e aplicá-los (Krüger, 2020). Costas (2017) apresenta um cenário mais abrangente, dizendo haver uma série de infraestruturas que subsidiam as práticas avaliativas com missões, características e focos específicos.

Assim como as bases de dados bibliográficas, os mecanismos de busca acadêmicos passam a constituir os importantes equipamentos informacionais para acesso e análise de publicações científicas atualizadas. Ambos se tornaram tão relevantes na atualidade que muitas das análises bibliométricas são baseadas em informações de publicações científicas obtidas a partir dessas fontes de informação (Gusenbauer, 2019).

Os buscadores se diferem das bases de dados porque são capazes de indexar fontes de informação diversas, assim como canais de comunicação formais e informais, incluindo *sites*, repositórios, redes sociais acadêmicas, entre outros, que se configuram como ferramentas úteis para a indexação de publicações científicas.

Vale frisar que, ao considerar os diversos indicadores de avaliação, deve-se atentar às informações presentes nas bases de dados e buscadores, como também compreender os cálculos que os caracterizam, pois seu valor final dependerá das características das informações consideradas nessas infraestruturas. Erros e manipulações nas fontes de dados e informações podem distorcer indicadores e demais estatísticas de impacto da produção científica.

As bases de dados e os buscadores acadêmicos

As bases de dados, de modo geral, visam ao gerenciamento das coleções, nas quais grandes quantidades de informações são organizadas e estruturadas de modo a possibilitar uma consulta facilitada a diversos documentos. Para isso, é necessário que essas bases de dados adotem critérios de padronização, de acordo com normas específicas, com o intuito de organizar e estruturar as produções bibliográficas.

Para Glänzel (2003), as fontes de informação mais confiáveis são as grandes bases de dados bibliográficas gerais ou multidisciplinares, que podem ser acessadas via internet. Essas bases de dados são referenciais, consistindo em um sistema que padroniza e sistematiza os conteúdos disponíveis nos mais diferentes meios informacionais, sendo fundamentais para as práticas da pesquisa científica. Assim sendo, as bases de dados tiveram – e ainda têm – um papel importante e homogeneizante nas práticas de editoração e de publicação eletrônica (Santos, 2011). Devido à sua relevância, elas também são instrumentos importantes para análises quantitativas da informação.

Dentre as bases de dados multidisciplinares mais importantes para a pesquisa cientométrica e bibliométrica, são mencionadas as seguintes: Web of Science (WoS), Scopus e Dimensions. Tradicionalmente, a avaliação via estudos cientométricos utiliza com mais assiduidade a WoS e/ou Scopus, sendo frequentemente comparadas quanto à cobertura e aos índices de citação. Conforme apontam Martín-Martín e demais autores (2021), nas muitas comparações que foram feitas entre essas bases, pode-se destacar que a Scopus é a mais compreensiva, enquanto a WoS é a mais seletiva.

Nesse sentido, Krüger e Petersohn (2022) destacam que, apesar dos esforços da WoS e da Scopus em diversificar as publicações indexadas, ampliando seu escopo para outros tipos de publicação – como anais de conferências ou livros, por exemplo – e outros idiomas além do inglês, também houve a diversificação dos tipos e das fontes de dados oriundas da digitalização da publicação acadêmica e da ascensão das mídias sociais, fora do universo fechado dos bancos de dados de citações.

Os autores Orduña-Malea e Delgado-López-Cózar (2018) consideraram o Dimensions como um produto potencial. De acordo com Bornmann (2018), uma característica interessante do Dimensions é o seu esquema

de classificação de campo, que é feito em aprendizado de máquina e não baseado em sistemas de classificação de periódicos, como ocorre na WoS e Scopus.

Singh e demais autores (2021) fizeram uma análise comparativa da cobertura de periódicos no Dimensions, na WoS e na Scopus. Nessa pesquisa, os autores notificaram que a cobertura de periódicos por essas bases difere significativamente. Eles consideram a WoS foi mais seletiva e o Dimensions foi mais exaustivo, aproveitando para pontuar outros aspectos importantes sobre sua cobertura.

Já entre os mecanismos de buscas acadêmicos, destacam-se na literatura especializada o Google Scholar (GS) e a Microsoft Academic (MA). Trata-se de infraestruturas que compõem um estoque de informação robusto de conhecimento científico, dos mais variados tipos documentais, pois são ferramentas capazes de filtrar as informações que são consideradas mais relevantes. Dentre estes, o GS é a fonte mais frequentemente utilizada e, segundo Gusenbauer (2019), isto pode estar relacionado à conveniência e à familiaridade dos usuários com o sistema do buscador generalista Google.

Para Martín-Martín e demais autores (2021), embora os buscadores acadêmicos sejam fontes de dados bibliográficos mais abrangentes, suas funcionalidades de pesquisa têm uma série de limitações dos recursos de buscas e filtragem, tais como a inconsistência dos resultados com uso de operadores booleanos, assim como o uso de uma série de algoritmos, cujas regras não são transparentes, comumente aplicados no processamento e classificação dos documentos na página de resultados.

A MA, assim como o GS, surgiu visando servir de mecanismo de busca gratuito de publicações científicas presentes em outras bases de dados na *web*. Costas (2017) afirmou que ambos apresentam limitações similares, pois não permitem estimar sua cobertura, além da baixa qualidade de metadados oferecidos. Outros estudos bibliométricos interessantes utilizaram a MA, tais como os de Harzing e Alakangas (2017), Hug e Brandle (2017) e Thelwall (2017, 2018a, 2018b).

Finalmente, o GS (ou Google Acadêmico) é uma versão de mecanismo de busca para produções científicas do Google Inc. – uma empresa especialista em soluções algorítmicas para lidar com a profusão informativa e,

atualmente, é um dos metabuscadores que mais faz parte da nossa vivência com os meios tecnológicos, sendo o principal mediador entre humanos e a informação na *web*.

Quadro 1 – Comparativo das características de WoS, Scopus, GS e Dimensions

Características	Web of Science	Scopus	Google Scholar	Dimensions
Lançamento	1997	2004	2004	2018
Mantenedor/ Desenvolvedor	Clarivate Analytics (desde 2016)	Elsevier	Google Inc.	Digital Science
Tipo	Base de dados	Base de dados	Buscador acadêmico	Base de dados
Resumo	Trata-se de uma plataforma de acesso para índices de citação multidisciplinares e regionais, índices de assuntos especializados, índice de patentes e índice para um conjunto de dados científicos.	É uma base de dados de resumos e citações da literatura científica revisada por pares, com ferramentas bibliométricas para acompanhar, analisar e visualizar a pesquisa.	É um mecanismo de busca virtual cuja finalidade é organizar e tornar público o conteúdo da literatura científica disposto em diversas outras plataformas da <i>web</i> .	É uma base de dados de bolsas de pesquisa, que vincula bolsas a publicações resultantes, ensaios clínicos e patentes. Fornece uma visão mais ampla de todo ciclo de vida da pesquisa, antes e além das publicações e citações.
Tamanho e cobertura	> 79 milhões (Core Collection) e 171 milhões de registros (plataforma); > 34 mil títulos de periódicos.	> 87 milhões de registros; > 25 mil títulos de periódicos.	> 389 milhões de registros (estimado em 2019); > Total de periódicos não estimado.	> 150 milhões de registros; > 93 mil títulos de periódicos.
Tipos de documentos	Artigos de periódicos, resumos, críticas, anais de conferências, livros, capítulos de livros, material editorial, artigo de acesso antecipado, cartas, relatórios técnicos, artigos de notícias, retratações, revisões, patentes.	Artigos de periódicos, anais de conferências, livros, periódicos, relatórios técnicos.	Artigos de periódicos, anais de conferências, livros, teses e dissertações, patentes, jurisprudência, <i>preprints</i> , relatórios técnicos e seus rascunhos, resumos, resenhas, editoriais etc.	Informações sobre bolsas e projetos de pesquisas, dados de financiamentos e financiadores, anais de conferências, <i>datasets</i> científicos, artigos científicos, <i>preprints</i> , livros, capítulos de livros, citações, patentes, ensaios clínicos, documentos de políticas governamentais entre outros dados altimétricos como menções em redes sociais etc.

Características	Web of Science	Scopus	Google Scholar	Dimensions
Cobertura de assuntos	Multidisciplinar (depende da seleção da base de dados).	Multidisciplinar.	Multidisciplinar.	Multidisciplinar.
Período de cobertura	1800 – presente (considerando todos os índices).	1788 – presente.	Sem período de cobertura para publicações <i>on-line</i> . O Scholar Metrics abrange artigos publicados entre 2017 e 2021.	1665 – presente.
Cobertura de idiomas	> 95% documentos em inglês.	> 92% documentos em inglês.	Informação não especificada, mas superior em outros idiomas quando comparado às demais bases de dados.	Informação não especificada.
Atualização	Cada coleção é atualizada em sua própria programação, variando de diária a mensal.	Diária.	Semanalmente/ mensalmente.	Diária para publicações. Dados de subsídios são atualizados mensalmente. Dados de patentes, semanalmente. Testes clínicos e <i>datasets</i> , quando solicitados.
Requer assinatura	Sim.	Sim.	Não.	Sim, mas possui uma versão gratuita.

Fonte: adaptado de Gusenbauer (2019); Orduña-Malea e Delgado-López-Cózar (2018); Vera-Baceta, Thelwall e Kousha (2019); Martín-Martín *et al.* (2021); Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019); Clarivate Analytics (2022).

No Quadro 1, verificam-se comparações das características dessas infraestruturas, nas quais se destacam suas diferenças e semelhanças. As publicações dos autores citados são algumas das fontes consideráveis para aqueles que procuram compreender ou observar as comparações práticas entre essas bases de dados e esses buscadores.

O que vale destacar é que as bases de dados e os buscadores acadêmicos citados compõem as diferentes fontes de informação, que possuem diversas funções que tornam possíveis o uso de ferramentas para análises

bibliométricas. Muitas das análises visam saber qual dessas plataformas tem o melhor desempenho sob perspectivas diversas. Entretanto, para se medir o impacto da pesquisa depende-se, inclusive, da área de pesquisa à qual está agregada e a sua cobertura.

Considerar que todos esses instrumentos servem exatamente para o mesmo fim e atribuir as vantagens e desvantagens de forma padronizada – para todas as áreas e seus propósitos – representará problema na avaliação de desempenho científico. A inexistência de uma base de dados ou buscador acadêmico que apresente uma cobertura completa da publicação científica mundial exige a seleção das fontes mais adequadas para as condições da análise desejada, incluindo abrangência geográfica, área do conhecimento e período temporal, por exemplo.

O GS

Esta seção busca aprofundar os conhecimentos sobre o GS. Serão descritas as origens, características, limitações, potencialidades do GS, como também serão observadas as métricas utilizadas pela ferramenta, que embasam suas diversas análises bibliométricas. Esta revisão bibliográfica é importante para descrever os conceitos que compreendem o uso do GS. A versão beta do buscador acadêmico surgiu em novembro de 2004. Em 2022, contamos com uma versão muito mais robusta do GS, da qual vamos saber mais detalhes a seguir.

Origens, objetivos e características

Nos últimos anos, outros buscadores surgiram e funcionam como mediadores de informação. De acordo com Ortega (2014), os buscadores acadêmicos podem ser considerados o ponto de encontro entre duas correntes que começaram a divergir com a evolução da *web*: I) as tradicionais bases de dados especializadas e II) os novos buscadores generalistas na *web*. Geralmente esses serviços de busca são gratuitos e baseados na *web* que podem ou não oferecer outras funcionalidades de valor agregado – tais como citações, indicadores etc.

O GS, desde o seu lançamento, foi considerado o primeiro buscador especializado em informações científicas dispostas na *web*, pois mostrou-se capaz de coletar e indexar automaticamente documentos cientificamente relevantes, escritos em qualquer idioma e publicados em qualquer país por meio do processo de rastreamento (Ortega, 2014).

Conforme destaca Cecchino (2010), os resultados da pesquisa feita no buscador podem incluir citações de literatura acadêmica, publicações revisadas por pares, teses, livros, resumos, artigos de editores acadêmicos, organizações profissionais, repositórios de pré-impressão, universidades e demais organizações acadêmicas, além de outras literaturas. Também oferece uma ferramenta de pesquisa avançada que ajuda parcialmente no controle da recuperação, pois não é tão especializado quanto as buscas robustas das bases de dados, haja vista que surgiu de um serviço de busca mais generalista (Ortega, 2014).

A razão pela qual passou a extrair e indexar os metadados de documentos científicos era simplesmente melhorar o posicionamento e ordenação dessa informação, considerada especialista, nos resultados da busca do próprio Google. No entanto, perceberam haver tipos diferentes de usuários do buscador: o usuário generalista – aquele que busca informações de um produto ou serviço, por exemplo – e o usuário acadêmico, estudante ou pesquisador – aquele que estava buscando conteúdo científico e cujas buscas eram mais estruturadas –, descreveram Orduña-Malea e demais autores (2016). A profusão de informações científicas disponíveis na internet e a descoberta do tipo de usuário interessado nelas foram o *start* que eles precisavam.

Inevitavelmente, a ferramenta impressiona, seja pela tecnologia empregada – mesma dinâmica usada no próprio buscador Google –, seja pela sua interface simples e fácil de usar, tornando o processo de busca por informações mais rápido e descomplicado, pois é seu objetivo oferecer acesso à literatura científica de forma mais dinâmica, indicando o estado atual da *web*, isto é, aquilo que é visível e rastreável para os seus robôs (*crawlers*). O próprio Google passou a fazer negociações com editoras científicas, universidades, bibliotecas eletrônicas e outras entidades para motivar a disponibilização de informações no GS, pois é precioso para eles o acesso ao PDF dos textos completos ou, pelo menos, o resumo daquela informação.

Dessa forma, tanto o Google passou a exigir algumas regras quanto aqueles que desejavam que seu conteúdo fosse indexado procuravam oferecê-los (Orduña-Malea *et al.*, 2016).

Ortega (2014) descreve como é feito o processo de rastreamento do GS. Para isso, ele menciona que, para extrair as informações ou documentos científicos dos *sites* de universidades, repositórios digitais e editoras científicas, é necessário: 1) que esses documentos sejam artigos acadêmicos; e 2) que apresentem um resumo legível. Outro fato interessante é que as informações básicas dos registros dos documentos científicos devem ser identificadas por metadados.

Em seus primeiros anos, percebe-se que o GS não nasceu com o intuito de ser um concorrente direto das bases de dados comerciais, tanto que Orduña-Malea e demais autores (2016) mencionam que ele não gera lucro para a empresa, sobretudo por não haver publicidade nos resultados das buscas geradas. Todavia, cabe ressaltar que se trata de um projeto ambicioso, pois, como lembra Gusenbauer (2019, p. 178), o objetivo dele é indexar “todo o universo das informações acadêmicas”. Esse fato acaba por fazer o escopo do GS um grande mistério. Apesar disso, faz com que o GS tenha uma cobertura superior às demais bases de dados (Orduña-Malea *et al.*, 2016).

O GS (2022) nos seus “Termos de isenção de responsabilidade” se esquivava de qualquer responsabilidade sobre a completude ou a precisão das informações indexadas. Diferentemente das bases de dados especialistas, muitas vezes não dita quais editoras, repositórios, universidades, agências de fomento, fundações e demais instituições de ensino e pesquisa compartilham seus documentos e nem quais documentos estão indexados. Em seu *site*, o GS também não fala como contabiliza as citações e com que frequência faz isso.

Apesar dessa obscuridade, nota-se que a ferramenta tem evoluído e melhorado em muitos aspectos, embora sua aceitação seja vista com cautela. Para os pesquisadores e profissionais da área de bibliometria e cientometria, o GS rendeu muitas observações e críticas. Peter Jacsó (University of Hawaii) é um dos pesquisadores mais conhecidos por fomentar estudos importantes e tecer duras críticas à ferramenta e àqueles que a defendem como totalmente apta para realização de estudos avaliativos.

Além disso, é preciso falar das essenciais contribuições de Orduña-Malea, Delgado-López-Cózar, Martín-Martín, Ayllón e demais colaboradores do EC3 Research Group (Universidad de Granada, Espanha) na publicação de estudos atuais do uso da ferramenta, artigos de revisão e livros sobre o GS, aspecto que também vale para o José Luis Ortega (Institute for Advanced Social Studies [IESA-CSIC], Espanha). Como também do empenho da professora Anne-Wil Harzing (Middlesex University Londres) que, desde 2006, auxilia bibliometristas do mundo todo com o *software* Publish or Perish (PoP).

Outrossim, cabe destacar como os estudos abordam o GS nos processos avaliativos. Percebe-se, assim como também é destacado por Orduña-Malea e demais autores (2016), que dentre os estudos que abordam a análise do GS como buscador acadêmico é possível destacar três frentes, em que há aqueles que: a) buscam conhecer a percepção da qualidade e da utilidade do produto por diferentes tipos de usuários – estudantes, acadêmicos e profissionais da informação; b) visam comparar os benefícios do GS com outros produtos de informação – catálogos, bases de dados bibliográficas, ferramentas de descoberta etc.; c) utilizam o GS como objeto de estudo, especificamente para analisar sua validade na avaliação da atividade acadêmica.

A primeira abordagem é a que mais se destaca perante as demais, visto que, dada a sua simplicidade e potencialidade, o buscador causou uma verdadeira revolução no comportamento dos usuários nos processos de busca e recuperação de informações científicas. Além da sua gratuidade e facilidade de uso, evidencia-se o reconhecimento da marca Google que, conforme destacam Costa, Canto e Pinto (2020), pode favorecer o uso desses recursos no mercado da avaliação científica.

Recursos

Genericamente, recursos são elementos que contribuem com algum tipo de benefício à sociedade. Para a realização de estudos avaliativos, é preciso saber as fontes e quais recursos informacionais existem, suas características, escopo, abrangência, usos, entre outros fatores determinantes. Logo, é de interesse da ciência da informação examinar os mecanismos que se dispõem

a funcionar como recursos de informação, justamente por ser convergente ao conceito de fontes de informação.

Basicamente, os produtos derivados ou secundários do GS são: o Google Scholar Metrics (GSM) e o Google Scholar Citations (GSC). O GS descreve o GSM como uma funcionalidade que oferece aos autores uma maneira simples de avaliar a visibilidade e a influência de seus artigos recentes. Ele resume as citações recentes de diversas publicações, auxiliando os autores na escolha do local para publicar suas novas pesquisas. Além disso, os usuários podem acessar as 100 principais publicações em vários idiomas, classificadas por suas métricas de índice h e mediana h de cinco anos. Ao clicar no número do índice h , é possível ver quais artigos foram mais citados em uma publicação específica, juntamente com as citações subjacentes às métricas (Google Scholar, 2022). Já o GSC se dirige aos autores, que monitoram as citações recebidas por seus artigos, além de poder analisar a literatura citante destes e de apresentar várias métricas de citação (Google Scholar, 2022).

Embora a descrição dada pelo próprio Google Scholar aos seus derivados seja simplista, percebe-se que esses recursos são e oferecem muito mais elementos do que apresentam. O Quadro 2 dispõe sucintamente o conteúdo de cada um deles.

Quadro 2 – Recursos secundários do Google Scholar: Google Scholar Citations e Google Scholar Metrics

	GSC	GSM
O que é	Lançado em 2011. Trata-se de um serviço gratuito que permite a geração de perfis de autores/pesquisadores – uma espécie de currículo – para exibirem uma lista de publicações que foram publicadas em diversas plataformas da <i>web</i> e é atualizado automaticamente.	Lançado em 2012. É um serviço de <i>ranking</i> de publicações científicas mais influentes.
Finalidade	a) Calcula as citações para esses artigos selecionados e implica indicadores bibliométricos, adicionados ao perfil; b) Os resultados da busca podem fundir versões do mesmo documento.	É um recurso útil para ajudar a localização de periódicos relevantes em uma especialidade ou país, exibindo diferentes classificações de acordo com indicadores bibliométricos.
Características	a) Contém informações pessoais estruturadas de formação – nome do pesquisador, afiliação e áreas de interesses de pesquisa; b) Tem como base os dados obtidos no GS, mas pode ser preenchido pelos próprios usuários que podem editar o conteúdo do seu próprio perfil; c) Oferece serviços adicionais, como alertas personalizados, listas de coautores, áreas de interesse, lista de autores por instituição etc.	O <i>ranking</i> de periódicos é baseado no índice h5, que se refere ao número h de um periódico, em que h artigos publicados nos últimos cinco anos tenham sido citados no mínimo h vezes cada. Também é calculada a mediana h5 dos artigos que compõem o índice h5.
Prós	a) Os usuários são os proprietários finais do perfil, editando-o com as informações que desejam, selecionando, mesclando e removendo publicações; b) Alto poder de divulgação das publicações científicas e outros trabalhos acadêmicos; c) Melhora a visibilidade e os impactos dos autores; d) É possível ver quem cita determinado autor ou documento; e) Permite criar alertas personalizados para se manter atualizado perante as áreas de interesse.	a) Ao contrário do resto dos serviços do GS, o GSM é uma fonte estática; b) A inclusão de diversas fontes documentais, algumas subvalorizadas, amplia o escopo da atividade de pesquisa e o alcance dessas publicações; c) Sua ampla cobertura engloba documentos de diversos idiomas, regiões e países e reforça a sua posição enquanto alternativa às bases de dados tradicionais.

	GSC	GSM
Contras	a) Possibilidade de manipulação intencional de perfis e adoção de práticas antiéticas; b) As edições realizadas pelos autores não ficam visíveis para os demais usuários; c) Atribuição errônea de papéis, ou na adição de pesquisadores de prestígio como coautores; d) Apresenta algumas inconsistências e limitações acerca da afiliação dos pesquisadores; e) Apresenta alguns erros nas contagens de citações; f) Impossibilidade de extração de dados para uso em estudos.	a) Inclui apenas periódicos que publicaram mais de 100 artigos nos últimos cinco anos – sendo uma restrição significativa para análises bibliométricas; b) Não é possível localizar prontamente os periódicos no GSM e, em seguida, acessar seus artigos no GS – que levanta a questão de qual é o valor real desta aplicação; c) Pelo GSM é bastante difícil saber exatamente quantos periódicos são indexados no buscador; d) O GSM difere do Journal Citation Reports (JCR) e Scientific Journal Rankings (SJR) pelas políticas de seleção, cobertura e indexação de periódicos científicos; e) Permite a inclusão de materiais diversos que, por outro lado, não passam por um processo de revisão por pares e a qualidade de seu conteúdo pode ser questionável ou algumas fontes são consideradas periódicos científicos, por exemplo, arXiv.org – repositório de <i>preprints</i> –, quando, na verdade, não são; f) Os únicos critérios de inclusão divulgados são que o periódico tenha ao menos 100 trabalhos publicados no período de avaliação e ao menos uma citação; g) A falta de normalização e padronização favorece a duplicação da entrada de alguns títulos; h) A mediana do índice h5 é uma medida incomum, o valor e a utilidade não são claros; i) Impossibilidade de extração de dados para uso em estudos.

Fonte: adaptado de Ortega (2014); Orduña-Malea *et al.* (2016); Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019); Costa, Canto e Pinto (2020).

Em 2011, Jacsó já mencionava que o GS se tornava mais robusto e mais inteligente com o passar do tempo, mas que ainda assim não era um recurso suficiente para ser usado em práticas bibliométricas ou para calcular uma variedade de indicadores informétricos de avaliação da quantidade ou qualidade da pesquisa científica. Nota-se que outros autores citados nessa pesquisa, com publicações mais atuais, ainda consideram que o GS não é uma ferramenta confiável, mas que pode ser utilizada como recurso complementar.

Ortega (2014) entende que os recursos do GS apresentam certa ingenuidade frente à complexidade que envolve a prática da bibliometria e dos estudos métricos da informação, tornando, inclusive, difícil de entender os propósitos desses recursos, uma vez que os diversos critérios adotados não são facilmente esclarecidos pelas informações dispostas na plataforma.

Obviamente ainda não existe nenhuma base de dados ou buscador acadêmico que seja perfeito a ponto de delinear com acurácia a qualidade e o impacto da ciência. Apenas uma combinação de indicadores bibliométricos diversos com processos, como a revisão por pares, por exemplo, pode compreender a natureza diversa do impacto científico de periódicos, pesquisadores ou grupos de pesquisa.

Tamanho, cobertura e crescimento

As bases de dados de citações são relevantes não somente por oferecer citações precisas, mas porque há uma qualidade muito estrita dos documentos que são indexados (Aguillo, 2012).

Em bases de dados como a WoS e Scopus, altamente controladas, é possível descobrir o número total de registros. Isso definirá o tamanho de determinada base de dados. Contudo, o GS é um motor de busca que apresenta o atual estado da *web* acadêmica. A *web* é um organismo dinâmico, instável e incontrolável. Portanto, eis o desafio metodológico de se medir o tamanho ou a quantidade de registros indexados pelo GS, uma vez que não é algo que pode ser verificado com precisão por conta das constantes variações. Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019) destacam que, no que diz respeito ao GS, para discutir o cálculo para estimar o seu tamanho é necessário entender as características da sua cobertura.

Sabe-se que o GS não fornece informação alguma sobre a sua cobertura, somente informações sobre os tipos de materiais que podem ser encontrados. Mesmo assim, não detalha a proporção de cada tipo documental nem quais são suas origens (Ortega, 2014). A partir do momento em que o GS rastreia, indexa ou exclui automaticamente algum documento científico sem nenhum controle ou auditoria prévia, o sistema nos leva à explosão de todos os controles ou filtros científicos aos quais foram submetidos.

Diversos estudos destacam que a cobertura do GS é superior quando comparada com outras bases de dados. O sistema PolySearch Engine, desenvolvido por Peter Jacsó (2005), visou estimar a cobertura do GS para determinadas editoras.

Estudos realizados por Mayr e Walter (2007), Walters (2007), Harzing (2014) e Harzing e Alakangas (2016) tentaram medir exploratoriamente a cobertura do buscador, seja comparando-o com outras bases de dados, seja na recuperação de publicações em diferentes áreas do conhecimento.

Para Jacsó (2012), a admiração exagerada ao GS vai ao encontro daqueles que desejam economizar recursos de bibliotecas, principalmente ao tentar transpor para o GS funcionalidades de recuperação da informação e outras funções que a ferramenta ainda não pode operar, fazendo com que suas métricas incertas passem a ser primordiais para avaliar o desempenho científico.

No entanto, Orduña-Malea e demais autores (2016) entendem que, com o passar dos anos, a cobertura do GS tem melhorado e se tornado mais dinâmica, superando as deficiências detectadas por diversos especialistas durante seus primeiros anos de atuação. Embora o GS permita uma boa cobertura de fontes, para se fazer análises mais aprofundadas – contando com indicadores bibliométricos mais sofisticados – é recomendada a utilização de outros *softwares* de apoio, sobretudo para sistematizar dados de citações (Aguillo, 2012; Mugnaini; Strehl, 2008).

Pode-se concluir que o GS é um sistema de *big data* acadêmico (Orduña-Malea *et al.*, 2016), pois inclui todos os tipos de documentos, de muitos outros idiomas e com uma taxa de crescimento superior tanto à da WoS e à da Scopus.

Como diz Harzing, no título do prefácio de Orduña-Malea e demais autores (2016), às vezes é preciso sacrificar um pouco de precisão por uma cobertura mais ampla. O contínuo crescimento da cobertura do GS surge à medida que ele coleta mais fontes acadêmicas abertas e faz mais acordos com as instituições ou participa da digitalização de antigas coleções. Contudo, esse aumento rápido leva alguns especialistas a criticar se o buscador visa à quantidade ao invés da qualidade da informação.

Erros e limitações do mecanismo

Conforme dito anteriormente, as publicações de Jacsó (2008a; 2008b, 2011) são as mais conhecidas por fomentar e tecer duras críticas ao GS, como também para aqueles que o defendem como mecanismo apto para realização de estudos avaliativos da informação científica.

Ortega (2014) aponta a opacidade dos resultados como um ponto crítico, em que se destacam as inconsistências dos filtros laterais dos resultados do GS. Outro ponto abordado pelo autor envolve o uso de operadores booleanos nas buscas que retornam resultados inexatos. Destaca que essas questões fomentam uma profunda desconfiança sobre a confiabilidade dos resultados do GS e, acima de tudo, sobre sua utilidade como um mecanismo de busca acadêmica.

Aguillo (2012) deixa claro que os problemas dos GS estão amplamente relacionados à falta de controle de qualidade e destaca que ele deve ser usado somente para fins complementares, não sendo adequado para fins bibliométricos ou usado como única ferramenta em qualquer tipo de avaliação.

Contudo, Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019) descrevem que, apesar da existência de muitas publicações sobre erros do GS, muitas delas não retratam o grau de abrangência dos erros nem definem como esses erros afetam os registros bibliográficos e suas citações. Para eles, faltam mais estudos empíricos. Os autores fizeram uma triagem dos erros comumente identificados no GS, compilados no Quadro 3, a seguir:

Quadro 3 – Tipos de erros comumente identificados no GS

Tipos	Descrição
Erros relacionados à pesquisa	São erros relacionados ao processo de busca de informações.
Erros relacionados à análise	São erros relacionados ao processo de identificação e extração de informações bibliográficas sobre documentos em <i>sítes</i> ou textos completos – incluindo referências citadas.
Erros de duplicações	São erros relacionados ao processo de identificação de diferentes versões de um mesmo documento, sobreposições e duplicatas.
Erros relacionados às fontes	São erros que afetam os <i>links</i> que levam à fonte na qual o documento foi encontrado.

Fonte: adaptado de Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019).

De qualquer forma, esses erros não devem ser ignorados, pois muitas vezes os dados oriundos do GS são usados sem verificação, o que pode trazer consequências graves e reais para muitas áreas de pesquisas, como as ciências da saúde. Em um estudo empírico mais recente realizado por Sauvayre (2022), verifica-se que uma das causas dos muitos erros do GS se deve à sua massiva indexação automática. A pesquisa abordou os erros do GS sob diferentes aspectos, dispostos no Quadro 4.

Quadro 4 – Erros do GS

Tipos	Descrição
Erros de coleta de dados	Inclui problemas relacionados às duplicatas, reimpressões, traduções, Localizadores Uniformes de Recursos (URLs) ausentes e documentos inacessíveis.
Erros de coleção de publicações acadêmicas	Apresenta os erros na identificação das tipologias documentais recuperadas e na inclusão de documentos não acadêmicos.
Erros de citação	Identifica documentos que não citaram a referência recuperada do GS, ou seja, as citações contadas pelo GS foram “falsos positivos”.
Erros no nome do autor	Equivale aos erros na descrição do nome do autor, muitas vezes advindo de iniciais removidas ou adicionadas, ou mesmo a substituição do nome do autor por outras informações das publicações. Também identificou a falta de alguns autores e até autores adicionados indevidamente em publicações.
Erros de título	Envolve aspectos como títulos incompletos, títulos semelhantes e outras informações indevidas adicionadas ao título.
Erros do ano de publicação	Identifica aspectos relativos à falta do dado ou substituição do dado por outros valores indevidos.
Erros de fonte	Apresenta as inconsistências de fonte, que ocorreram principalmente em nomes de periódicos, como erros tipográficos e erros de capitalização. Também envolve questões de incompletude do título e do nome do periódico, como também de nomes dos periódicos que foram fortemente truncados.
Erros do editor	Demonstra que editores de periódicos e nomes de domínio eram frequentemente imprecisos, incluindo o nome de editores substituídos por outras informações irrelevantes.

Fonte: adaptado de Sauvayre (2022).

Sauvayre (2022) conclui que quase todos os dados recuperados do GS em seu estudo continham, pelo menos, um tipo de erro. Então, destaca que, sem a devida limpeza dos dados, a confiabilidade do GS é altamente questionável, pois apontam erros que podem ter consequências importantes, como erros na contagem de citações. Um ponto interessante frisado se refere ao tempo despendido na verificação meticulosa das amostras oriundas do GS,

o que pode tornar o trabalho de análise ainda mais complexo, mesmo se utilizando de ferramentas complementares, como o *software* PoP.

Os erros e limitações do GS são muitas vezes observados conjuntamente, percebendo uma linha tênue entre as questões abordadas. Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019) são os autores que descrevem as limitações do GS com bastante detalhes. Os autores mencionam que essas limitações ocorrem nas seguintes instâncias: cobertura, interface de pesquisa e resultados, qualidade dos dados e capacidade de reutilização e exportação de dados.

A questão da cobertura do GS já foi descrita anteriormente. Todavia, os autores acrescentam outros pontos importantes:

- a falta de transparência da cobertura, de uma lista pública das fontes e de todos os periódicos indexados pelo GS;
- a inclusão de documentos não científicos e não acadêmicos nos resultados do GS;
- a falta de um método preciso para se estimar o tamanho do GS;
- a falta de estabilidade dos dados, o que pode enviesar algumas análises bibliométricas que não levarem isso em consideração;
- a permissividade para manipulação.

Quanto à interface de pesquisa e resultados, Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019) dispõem que:

- os resultados são hierarquizados, segundo a quantidade de citações que recebem, logo, os registros nas primeiras colocações serão os mais citados;
- a pesquisa avançada é muito limitada;
- a visão dos registros é limitada somente a dez por vez, diferentemente de outras bases de dados que permitem a visualização em mais dezenas;
- os filtros e a ordenação dos resultados são muito limitados;
- faltam recursos para analisar ou “computar” os indicadores bibliométricos.

No que se refere à qualidade dos dados, os autores (Delgado-López-Cózar; Orduña-Malea; Martín-Martín, 2019, p. 121) verificam que:

- muitos dados de autorias são suprimidos devido aos poucos caracteres reservados para essa descrição, dividindo o mesmo espaço com outros dados, tais como o nome do periódico, a fonte e o ano de publicação;
- não há dados disponíveis sobre a afiliação dos autores (instituição ou país);
- a tipologia de cada documento não é clara;
- nem todos os documentos apresentam um resumo;
- a lista de referências citadas em cada artigo também não está disponível;
- a descrição bibliográfica dos documentos é deficiente.

Por fim, no que tange à sua capacidade de reutilização e exportação de dados, Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019, p. 121) acrescentam que no GS:

- os registros só podem ser salvos manualmente, um por um;
- somente usuários logados, isto é, com um “cadastro” no Google têm acesso à função “minha biblioteca” do GS e é por meio desta que se tem acesso aos recursos de exportação das citações em .CSV ou para os gerenciados de referência;
- o resumo nunca é incluído nos registros exportados;
- o GS não oferece nenhum tipo de interface de APIs (interface de programação de aplicações).

Adiciona-se a isso o fato de o GS indicar as fontes disponíveis, mas, muitas vezes, não garantir o acesso a elas. O entendimento do que é um documento científico também precisa ser verificado pelo GS, uma vez que o conteúdo científico é disposto de maneira diferente em diversas áreas, como também em diferentes países.

Métricas do GS

Cabezas-Clavijo e Delgado-López-Cózar (2013) advertem que as novas ferramentas de acesso científico gratuito, como o GS, popularizaram o acesso à avaliação científica e, conseqüentemente, ao índice h – indicador bibliométrico padrão adotado por esses serviços. Os indicadores bibliométricos calculados pelo GS e seus recursos secundários são os índices h, i10, h5, h5 Core e a mediana h5. O GSC utiliza os índices h e i10 para mensurar as citações – métricas de nível de autor. Já o GSM utiliza os índices h5, h5 Core e mediana h5 para fins de avaliação da produção científica – métricas em nível de impacto de periódico.

De modo geral, conforme destacam Cabezas-Clavijo e Delgado-López-Cózar (2013), o índice h foi criado em um contexto de busca por indicadores válidos para medir o desempenho individual de forma mais objetiva. Basicamente, o índice h visa à convergência das dimensões quantitativas (produção) e qualitativas (citações) da pesquisa em um único indicador.

Conseqüentemente, especialistas em bibliometria começaram a discutir as diversas aplicações do índice h, seus prós e contras. Jacsó (2008b) fala que o índice h não deve ser aceito como um indicador único de desempenho.

De qualquer forma, percebe-se que, apesar das suas falhas, o uso do índice h se tornou generalizado. Seu cálculo pode ser feito direta e automaticamente pelas bases de dados, tornando a sua difusão mais rápida e compreensível, resultado do “populismo avaliativo”, destaca Gingras (2016).

Quadro 5 – GSC – métricas de nível de autor: índices h e i10

Métricas	Geral	Últimos cinco anos
Citações	Número de citações que um autor recebeu.	Número de citações que um autor recebeu nos últimos cinco anos.
Índice h	Número de artigos h que receberam pelo menos h citações.	Maior número h, tal que h publicações tenham pelo menos h novas citações nos últimos cinco anos.
Índice i10	Número de publicações com pelo menos dez citações, isto é, apenas documentos científicos que receberam dez ou mais citações, são considerados.	Número de publicações com pelo menos dez citações nos últimos cinco anos.

Fonte: adaptado de Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019).

Sobre os índices h e i_{10} utilizados no GSC, Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019) observam que esses indicadores são calculados considerando dois diferentes intervalos de tempo: primeiramente, o índice h não possui nenhuma restrição de tempo – útil para comparações de acadêmicos seniores –, em segundo momento, registram-se apenas as citações recebidas nos últimos cinco anos – sendo útil para comparações de pesquisadores de carreira. Reforçam ser importante ter em mente que esses indicadores são calculados automaticamente a partir dos dados na lista de publicação, sem qualquer tipo de supervisão.

Por outro lado, a avaliação da produção científica em diferentes áreas, instituições e países se tornou um instrumento necessário para mapear a ciência, contribuindo para a tomada de decisões de políticas científicas. Como posto na justificativa, a avaliação da qualidade dos periódicos vem sendo amplamente empregada no Brasil, até mesmo como forma de análise de programas de pós-graduação e fazendo parte de critérios do Qualis da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), destacam Thomaz, Assad e Moreira (2011).

Para avaliação dos periódicos científicos, o Fator de Impacto (FI) é comumente empregado. Mesmo com o seu prestígio, sabe-se que os periódicos de países como o Brasil têm menor visibilidade internacional e baixo valor do indicador, como alerta a Sociedade Brasileira de Medicina Tropical (2020).

Como alternativa ao FI, o índice h surge como uma forma de avaliar o impacto dos periódicos, tendo como base as informações de citações. O emprego do índice h para avaliação do impacto científico advém da própria característica do índice: trata-se de uma medida fácil de calcular e se baseia em critérios objetivos – combina produtividade e impacto científico em um único indicador, destacam Silva, Almeida e Grácio (2018) e Costas e Bordons (2007).

O GS optou pela utilização do índice h , talvez por sua simplicidade e robustez, apesar das suas conhecidas limitações e inconsistências. O índice h não é calculado para um período fixo específico como o FI, mas pode ser usado para calcular período desejado, por exemplo, por um ou cinco anos. Além disso, incluiu a possibilidade de ordenação pela mediana do índice h –

mediana do número de citações que contribuem para o índice h calculado (Orduña-Malea *et al.*, 2016).

No GSM, os periódicos são categorizados por idioma. Ao selecionar um idioma, o sistema exibe uma lista dos 100 periódicos com o maior índice h. Ao clicar no índice h correspondente, é exibido o conjunto de artigos que contribuíram para aquele valor, ordenados pelo número de citações recebidas por artigo. Por fim, clicando no título de cada artigo, o texto completo pode ser acessado – se disponível.

Costas e Bordons (2007) apontam que tanto a quantidade quanto o impacto das publicações são considerados no cálculo do índice h, mas que o número de publicações desempenha um papel muito importante, pois é o índice h máximo que um autor pode obter.

Quadro 6 – GSM – métricas para avaliação de periódicos: índices h5, h5 Core e mediana h5

Métricas	Descrição
Índice h5	O maior número h tal que pelo menos h artigos nesse periódico foram citados pelo menos h vezes cada um nos últimos cinco anos.
Índice h5 Core	O conjunto de artigos de um periódico com uma citação acima do limite do índice h5.
Mediana h5	Mediana da distribuição de citações aos artigos no h5 Core.

Fonte: adaptado de Delgado-López-Cózar, Orduña-Malea e Martín-Martín (2019).

Percebe-se o quanto o número de publicações é importante em Vargas, Vanz e Stumpf (2014), que ressaltaram que o aumento da indexação em bases de dados e na quantidade de publicações da área de ciências agrárias ajudou a elevar o Brasil no *ranking* global de produção científica. Os autores também observaram que, embora mais periódicos tenham sido indexados na WoS, houve uma mudança significativa na comunicação científica da área, em que também se aumentou o número de fascículos anuais, assim como o número de artigos, resultando, consequentemente, em um maior índice h, pois os valores aplicados aos periódicos bem avaliados são transferidos para seus autores.

Ocorre que, independentemente do uso para avaliação do pesquisador ou do impacto das publicações, o índice h deve ser observado com

cautela. Em muitas disciplinas, o índice não representará o impacto real, pois cada área tem suas peculiaridades e diferentes tendências de citação. Acrescenta-se o fato do impacto de uma publicação não se medir apenas por publicações, mas por sua contribuição para inovações tecnológicas, por exemplo (Marques, 2013).

Vale ressaltar que o desempenho em pesquisa é um empreendimento complexo e multifacetado que não pode ser avaliado adequadamente por meio de um único indicador, como também não deve ser o elemento crucial para decisões de políticas científicas ou como único critério de avaliação para medir o desempenho de grupos de pesquisadores, universidades, países e periódicos. Todo e qualquer índice bibliométrico terá suas vantagens e limitações. Para uma melhor prática avaliativa e para evitar perigosos desvios, a utilização de um conjunto de indicadores representa a forma mais justa e legítima para ir além das generalidades impostas por bases de dados e buscadores acadêmicos.

GS e a visibilidade da produção científica

O GS permite que parte oculta da ciência seja visível, especialmente em áreas das ciências sociais e humanidades, assim como temas científicos regionais ou oriundos de países periféricos. Orduña-Malea e demais autores (2016) destacam que o GS é a porta de entrada, pois permite o acesso gratuito às informações científicas para os usuários que, porventura, não estão vinculados a alguma instituição que pagam o acesso às bases de dados comerciais como WoS e Scopus. Outro fator que influencia na sua visibilidade é a cobertura ampla em termos de países, idiomas e tipologias documentais para além do periódico científico, algo que é importante para determinadas áreas que usufruem de diferentes canais de comunicação (Delgado-López-Cózar; Orduña-Malea; Martín-Martín, 2019).

De acordo com Harzing (2016b), a cobertura do GS melhora com o passar dos anos, sendo, muitas vezes, superior à da WoS ou à da Scopus. A autora já havia mencionado isso em outros estudos, tais como Harzing e Mijnhardt (2015) e Harzing e Alakangas (2016), que frisam que o uso do GS em estudos bibliométricos se tornou tangível após a concepção do

PoP. Além disso, menciona a cobertura superior do GS em áreas de ciências sociais e humanidades e outras áreas cuja comunicação científica ocorre por meio de livros, congressos e conferências.

Orduña-Malea e demais autores (2016) também reforçam esse ponto e dizem que para algumas áreas – como computação, eletrônica, informática e telecomunicações, por exemplo – os eventos científicos são essenciais instrumentos de divulgação científica, pois são áreas em que a comunicação é mais dinâmica. Em 2019 e em 2020, no *ranking* dos artigos mais influentes e mais citados no GS, divulgados pelo periódico *Nature*, muitos são sobre tecnologia, aprendizado de máquina e inteligência artificial. Em 2021, apesar de os artigos sobre a covid-19 terem alcançado o protagonismo, percebe-se que esses temas continuam chamando a atenção. Embora as diversas bases de dados possuam publicações resultantes de eventos científicos em seu escopo, notam-se diferentes graus de exaustividade, destacam os autores (Orduña-Malea *et al.*, 2016).

Harzing (2016b) também destaca que a cobertura geográfica e linguística do GS é interessante, pois indexa documentos com idiomas além do inglês, fato que já tinha o reconhecimento de Jacsó (2008a).

Harzing (2016b) também descreve que, embora o GS apresente problemas críticos, o buscador também desempenha um importante papel na democratização da análise de citações, embora tais análises só sejam possíveis mediante o uso dos *softwares* de apoio como o PoP. A autora ressalta que se for necessária uma precisão de 99-100%, o GS sempre precisará ser triangulado com outras fontes de dados.

Embora a tensão latente acerca do buscador ainda ocorra e suas limitações há tempos sejam debatidas na literatura especializada, sobretudo para o uso do GS para fins avaliativos, as análises bibliométricas do GS feitas por Martín-Martín e demais autores (2018) também constataram que o GS tem grande importância quanto à visibilidade. Por isso, permite complementar as outras bases de dados especialistas em estudos bibliométricos. A visibilidade da produção em ciência aberta é outro principal atrativo para ter os trabalhos indexados no GS, conforme apontam Torres-Salinas, Ruiz-Pérez e Delgado-Lopez-Cózar (2009).

Canto e demais autores (2022) mencionam que a utilização do GS e seus recursos são frequentemente mencionados em estudos como ferramenta alternativa para análise bibliométrica nas áreas de artes, humanidades e ciências sociais, inclusive os publicados exclusivamente em português e espanhol.

Para Jacsó (2008a), a cobertura de repositórios digitais é uma das vantagens do GS. Sobre esse ponto, no caso da América Latina e do Caribe em particular, o GS desempenha um papel importante, pois, devido à sua ampla cobertura, é capaz de acessar e analisar o impacto dos periódicos nesses países, que estão muitas vezes disponíveis em repositórios institucionais, portais de periódicos, *sites* e bibliotecas digitais como a Scientific Eletronic Library Online (SciELO).

Como apontam Canto e demais autores (2022), o GSM é uma fonte relevante para a América Latina e o Caribe, pois a ampla cobertura permite a análise de impacto para um número maior de periódicos. Segundo os autores, os periódicos oriundos do Brasil, do México, da Colômbia, da Argentina, do Chile e da Cuba correspondem a quase 90% do conjunto de publicações indexadas. Com isso, nota-se que o GS é uma ferramenta útil para divulgação e impactos de periódicos de uma região que nunca teve destaque no campo científico nas bases de dados internacionais, por refletir uma tendência temática regional. Embora muitas vezes estejam preocupados com questões atuais que afetam áreas específicas e escritos em idiomas regionais, agora eles têm um modelo de publicação de acesso aberto e seu impacto pode ser medido, considerando as fontes regionais (Santos *et al.*, 2021).

Tem-se observado o crescente uso do GS em estudos altmétricos, observam Orduña-Malea e demais autores (2016), que sugestionam que deles virão grandes tendências no uso da ferramenta para análises bibliométricas.

Ao fazer uma busca do próprio GS com os termos “visibilidade”, “produção científica” e “google scholar” ou “google acadêmico”, por exemplo, recupera-se cerca de 560 documentos, 523 deles em língua portuguesa, sendo sua grande maioria oriundos do Brasil. Somente na primeira face da Página de Resultados de Pesquisa dos Buscadores (Serp) encontram-se muitos trabalhos que discutem a visibilidade que o GS proporciona para produções intelectuais de repositórios institucionais universitários, teses

e dissertações, humanidades digitais e demais áreas das ciências sociais e humanidades. Não significa que os trabalhos não discutam criticamente a capacidade do GS enquanto ferramenta para fins avaliativos. Autores como Mugnaini e Strehl (2008), Bueno e Caregnato (2013), Silva e Grácio (2017) aparecem na primeira página. Evidencia-se, para estudos posteriores, a importância de avaliação desse conteúdo que pode trazer discussões importantes para essa questão.

Considerações finais

O capítulo buscou analisar a relevância das fontes de informação, base essencial das infraestruturas para a realização dos estudos bibliométricos de qualquer natureza. Verificou-se que entre as principais fontes de informação estão as bases de dados e os buscadores acadêmicos. Entre os buscadores acadêmicos, o GS se destaca por ser uma fonte disponível gratuitamente na internet, que organiza e torna público o conteúdo da literatura científica dispostos em diversas plataformas da *web*. Esse conteúdo vai além daqueles publicados em periódicos, fato que torna o tamanho e a cobertura do GS muito maiores que os das bases de dados tradicionais.

Percebe-se que o GS chama a atenção da comunidade de bibliometria e cientometria devido aos seus recursos secundários: o GSM e o GSC que, respectivamente, permitem a obtenção de algumas métricas de avaliação da visibilidade e da influência de publicações acadêmicas, como a geração de perfis de pesquisadores e o cálculo da citação de artigos. Contudo, devido à complexidade que envolve as análises bibliométricas, ter esses recursos à disposição sem a devida transparência dos critérios adotados pela plataforma dá margem para discussões e críticas feitas pela comunidade de bibliometria e cientometria.

Portanto, é possível afirmar que a comunidade de bibliometria e cientometria deve acompanhar a evolução e os aprimoramentos do GS e seus recursos, pois constantemente utiliza o buscador em seus estudos como ferramenta alternativa. Contudo, cabe destacar que seu uso deve ser feito com o devido cuidado, ciente das suas deficiências e das consequências do seu amplo uso para avaliar e mensurar o desempenho científico.

De modo geral, pode-se destacar alguns aspectos importantes sobre o uso do GS para fins avaliativos:

- sobre suas potencialidades: devido ao tamanho e à diversidade documental, percebe-se que o buscador se tornou uma ferramenta útil e importante para a visibilidade de informações científicas contidas em outras fontes de informação, diferente das contidas nas tradicionais bases de dados, tais como repositórios, *sites* institucionais e pessoais, publicações em acesso aberto etc. Pelo GS também é possível ter acesso às produções científicas de áreas cujo trabalho não tem o periódico como canal majoritário de publicação. Ademais, permite que outros tipos de material possam ser analisados por meio de citações. Por fim, também torna acessível a literatura científica de países com outros idiomas, além do inglês, e de nações em desenvolvimento, que muitas vezes tratam de temas mais regionais.
- sobre os erros e limites do GS: percebe-se que os autores, de modo geral, relatam as várias limitações das ferramentas, muitas delas ligadas à opacidade do buscador e à falta de transparência sobre os critérios adotados. Os erros, muitas vezes, são oriundos dos metadados das publicações, que diferem das informações verídicas. Além disso, vale destacar que o PoP se tornou um *software* auxiliar importante para a manipulação e análise dos dados obtidos no GS, ainda que muitos autores considerem seu uso impróprio, quando não se empreende o devido trabalho de limpeza dos dados. Assim, esses erros não devem ser ignorados e a não verificação de seus problemas pode trazer consequências graves para muitas áreas, ao tornar os processos avaliativos injustos.

Pode-se concluir pelas peculiaridades observadas nesses estudos e, em particular, pelas limitações descritas por vários dos autores, que nesta publicação são oferecidas orientações para: a) aqueles que planejam seus próprios procedimentos metodológicos; b) aqueles que recorrem aos serviços de métricas gerados por ferramentas diversas; c) os exercícios de avaliação que utilizam a GS e os indicadores resultantes dele para a determinação de critérios objetivos. Como exemplo desse último item, pode-se ressaltar o “Novo Qualis” da Capes, no qual a inclusão do GS para a determinação

dos estratos dos periódicos deve ser observada com cautela, tendo critérios transparentes.

Ressalta-se, ainda, que este capítulo traz diversas referências e contribuições importantes da comunidade da área de bibliometria e cientometria para entendimento do tema.

Referências

AGUILLO, I. F. Is Google Scholar useful for bibliometrics? A webometric analysis. *Scientometrics*, [Budapest], v. 91, p. 343-351, Dez. 2012. DOI 10.1007/s11192-011-0582-8. Disponível em: <https://link.springer.com/article/10.1007/s11192-011-0582-8>. Acesso em: 15 out. 2020.

BORNMANN, L. Field classification of publications in Dimensions: a first case study testing its reliability and validity. *Scientometrics*, [Budapest], v. 117, p. 637-640, 2018. DOI 10.1007/s11192-018-2855-y. Disponível em: <https://link.springer.com/article/10.1007/s11192-018-2855-y>. Acesso em: 15 mar. 2021.

BUENO, L. S.; CAREGNATO, S. E. Citações em rede: a visibilidade do texto acadêmico na web. In: SALÃO INICIAÇÃO CIENTÍFICA, 25., Anais [...], Rio Grande do Sul, UFRGS, 2013. Disponível em: https://www.lume.ufrgs.br/bitstream/handle/10183/91343/Poster_29092.pdf?sequence=2. Acesso em: 5 mar. 2021.

CABEZAS-CLAVIJO, A.; DELGADO-LÓPEZ-CÓZAR, E. Google Scholar and the h-index in biomedicine: The popularization of bibliometric assessment. *Medicina Intensiva*, España, v. 37, n. 5, p. 343-354, jun./jul. 2013. Disponível em: https://www.sciencedirect.com/science/article/pii/S2173572713000398?casa_token=MtXpUlvZRFwAAAAA:KEwfrw_z3xQ4cUuSAs_OHCJUs2hrRIF8UJX3O5G2XaDeTJqs26zHm_V9Kew1LidWnlrzgNuUzw. Acesso em: 2 ago. 2021.

CANTO, F. L. *et al.* Latin American and Caribbean journals indexed in Google Scholar Metrics. *Scientometrics*, [Budapest], v. 127, n. 2, p. 763-783, Feb. 2022. DOI 10.1007/s11192-021-04237-x. Disponível em: <https://link.springer.com/article/10.1007/s11192-021-04237-x>. Acesso em: 2 mar. 2022.

CECCHINO, N. J. Google Scholar. *Journal of the Medical Library Association*, Chicago, v. 98, n. 4, p. 320-321, Oct. 2010. DOI 10.3163/1536-

5050.98.4.016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2947134/pdf/mlab-98-04-320.pdf>. Acesso em: 12 abr. 2022.

CLARIVATE Analytics. *Web of Science*, A História do ISI e o trabalho de Eugene Garfield. [S. l.], 2022. Disponível em: <https://clarivate.com/the-institute-for-scientific-information/history-of-isi/>. Acesso em: 15 jan. 2022.

COSTA, H.; CANTO, F. L. do; PINTO, A. L. Google Scholar Metrics e a proposta do novo Qualis: impacto dos periódicos brasileiros de Ciência da Informação. *Informação & Sociedade: Estudos*, João Pessoa, v. 30, n. 1, p. 1-16, 2020. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/50676/29814>. Acesso em: 18 jun. 2021.

COSTAS, R. Discussões gerais sobre as características mais relevantes de infraestruturas de pesquisa para a Cientometria. In: MUGNAINI, R.; FUJINO, A.; KOBASHI, N. Y. (org.). *Bibliometria e cientometria no Brasil: infraestrutura para avaliação da pesquisa científica na Era do Big Data*. São Paulo: ECA/USP, 2017. E-book. p. 19-66. DOI 10.11606/9788572051705. Disponível em: <https://www.livrosabertos.abcd.usp.br/portaldelivrosUSP/catalog/view/129/108/547>. Acesso em: 16 jan. 2021.

COSTAS, R.; BORDONS, M. The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *Journal of Informetrics*, [s. l.], v. 1, n. 3, p. 193-203, June. 2007.

DELGADO-LÓPEZ-CÓZAR, E.; ORDUÑA-MALEA, E.; MARTÍN-MARTÍN, A. Google Scholar as a data source for research assessment. In: GLÄNZEL, W. et al. (ed.). *Springer Handbook of Science and Technology Indicators*. New York City: Springer, 2019. p. 96-126. Disponível em: https://www.researchgate.net/publication/325733835_Google_Scholar_as_a_data_source_for_research_assessment. Acesso em: 15 jul. 2021.

GINGRAS, Y. *Os desvios da avaliação da pesquisa: o bom uso da bibliometria*. Rio de Janeiro: UFRJ, 2016.

GLÄNZEL, W. *Bibliometrics as a research field: a course on theory and application of bibliometric indicators*. [Budapest: s. n.], 2003. Disponível em: https://www.cin.ufpe.br/~ajhol/futuro/references/01%23_Bibliometrics_Module_KUL_BIBLIOMETRICS%20AS%20A%20RESEARCH%20FIELD.pdf. Acesso em: 15 dez. 2020.

GOOGLE SCHOLAR. *About*. [S. l.], 2022. Disponível em: <https://scholar.google.com/intl/en/scholar/about.html>. Acesso em: 12 fev. 2022.

GUSENBAUER, M. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*, [Budapest], v. 118, p. 177-214, 2019. DOI 10.1007/s11192-018-2958-5. Disponível em: <https://link.springer.com/article/10.1007/s11192-018-2958-5>. Acesso em: 15 jun. 2020.

HARZING, A.-W. A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics*, [Budapest], v. 98, p. 565-575, 2014. DOI 10.1007/s11192-013-0975-y. Disponível em: <https://link.springer.com/article/10.1007/s11192-013-0975-y>. Acesso em: 15 fev. 2021.

HARZING, A.-W. Microsoft Academic (Search): a Phoenix arisen from the ashes? *Scientometrics*, [Budapest], v. 108, p. 1637-1647, 2016a. Disponível em: <https://repository.mdx.ac.uk/download/ff4f0c3a092a74961b5854364f3f64d53bbdcd0463152a81e40a4d94a10de7b6/642328/mas.pdf>. Acesso em: 15 jan. 2021.

HARZING, A.-W. Sacrifice a little accuracy for a lot more comprehensive coverage. In: ORDUÑA-MALEA, E. et al. *La revolución Google Scholar: destapando la caja de Pandora académica*. Madrid: Unión de Editoriales Universitarias Españolas, 2016b. p. 15-20.

HARZING, A.-W.; ALAKANGAS, S. Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, [Budapest], v. 106, p. 787-804, 2016. DOI 10.1007/s11192-015-1798-9. Disponível em: <https://link.springer.com/article/10.1007/s11192-015-1798-9>. Acesso em: 15 jan. 2021.

HARZING, A.-W.; ALAKANGAS, S. Microsoft Academic: is the phoenix getting wings? *Scientometrics*, [Budapest], v. 110, p. 371-383, 2017. Disponível em: <https://link.springer.com/article/10.1007/s11192-016-2185-x>. Acesso em: 15 jan. 2021.

HARZING, A.-W.; MIJNHARDT, W. Proof over promise: Towards a more inclusive ranking of Dutch academics in Economics & Business. *Scientometrics*, [Budapest], v. 102, n. 1, p. 727-749, 2015. Disponível em: <https://harzing.com/download/top40.pdf>. Acesso em: 15 fev. 2021.

HUG, S. E.; BRÄNDLE, M. P. The coverage of Microsoft Academic: Analyzing the publication output of a university. *Scientometrics*, [Budapest], v. 113, n. 3, p. 1551-1571, 2017. DOI 10.1007/s11192-017-2535-3. Disponível em: <https://link.springer.com/article/10.1007/s11192-017-2535-3>. Acesso em: 15 fev. 2021.

JACSÓ, P. Google Scholar duped and deduped: the aura of “robometrics”. *Online Information Review*, [Bradford], v. 35, n. 1, p. 154-160, 2011.

JACSÓ, P. Google Scholar Metrics for publications: the software and content features of a new open access bibliometric service. *Online Information Review*, [Bradford], v. 36, n. 4, p. 604-619, 2012.

JACSÓ, P. Google Scholar revisited. *Online Information Review*, [Bradford], v. 32, n. 1, p. 102-114, 2008a. DOI 10.1108/14684520810866010. Disponível em: <https://tefkos.comminfo.rutgers.edu/Courses/e530/Readings/Jaszo%20Google%20scholar%202008.pdf>. Acesso em: 15 jun. 2020.

JACSÓ, P. The pros and cons of computing the h-index using Google Scholar. *Online Information Review*, [Bradford], v. 32, n. 3, p. 437-452, 2008b. DOI 10.1108/14684520810889718. Disponível em: <https://sci2s.ugr.es/sites/default/files/files/TematicWebSites/hindex/Jacso2008a.pdf>. Acesso em: ago. 2021.

JACSÓ, P. Visualizing overlap and rank differences among web-wide search engines: Some free tools and services. *Online Information Review*, [Bradford], v. 29, n. 5, p. 554-560, 2005.

KRÜGER, A. K. Quantification 2.0? Bibliometric Infrastructures in Academic Evaluation. *Politics and Governance*, [s. l.], v. 8, n. 8, p. 5867, 2020. DOI 10.17645/pag.v8i2.2575. Disponível em: <https://www.cogitatiopress.com/politicsandgovernance/article/view/2575/1490>. Acesso em: 15 fev. 2021.

KRÜGER, A. K. PETERSOHN, S. ‘I want to be able to do what I know the tools will allow us to do’: Practicing evaluative bibliometrics through digital infrastructure. *Research Evaluation*, Guildford, p. 1-11, 2022. Special issue. DOI 10.1093/reseval/rvac009. Disponível em: <https://academic.oup.com/rev/article/31/4/475/6749618>. Acesso em: 15 out. 2022.

MARQUES, F. Os limites do índice-h: Supervalorização do indicador que combina quantidade e qualidade da produção científica gera

controvérsia. *Revista Pesquisa FAPESP*, São Paulo, v. 207, p. 35-39, maio 2013. Disponível em: https://revistapesquisa.fapesp.br/wp-content/uploads/2013/05/035-039_IndiceH_207.pdf. Acesso em: 25 maio 2021.

MARTÍN-MARTÍN, A. *et al.* Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations. *Scientometrics*, [Budapest], v. 126, p. 871-906, 2021. DOI 10.1007/s11192-020-03690-4. Disponível em: <https://link.springer.com/article/10.1007/s11192-020-03690-4>. Acesso em: 15 mar. 2021.

MARTÍN-MARTÍN, A. *et al.* Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, [s. l.], v. 12, n. 4, p. 1160-1177, Nov. 2018.

MAYR, P.; WALTER, A. K. An exploratory study of Google Scholar. *Online Information Review*, [Bradford], v. 31, n. 6, p. 814-830, 2007. Disponível em: <https://arxiv.org/pdf/0707.3575>. Acesso em: 15 dez. 2021.

MUGNAINI, R. Produção e impacto da pesquisa brasileira: confrontando contextos internacional e nacional. *SciELO em Perspectiva*, São Paulo, 27 ago. 2018. Disponível em: <https://blog.scielo.org/blog/2018/08/27/producao-e-impacto-da-pesquisa-brasileira-confrontando-contextos-internacional-e-nacional/#.YJGrhLVKg2w>. Acesso em: 15 mar. 2021.

MUGNAINI, R.; STREHL, L. Recuperação e impacto da produção científica na Era Google: uma análise comparativa entre o Google Acadêmico e a Web of Science. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*. Florianópolis, p. 92-105, 1 semestre 2008. Número especial. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13nesp1p92/1570>. Acesso em: 15 jun. 2020.

ORDUÑA-MALEA, E.; DELGADO-LÓPEZ-CÓZAR, E. Dimensions: re-discovering the ecosystem of scientific information. *El Profesional de la Información*, Barcelona, v. 27, n. 2, p. 420-431, 2018. DOI 10.3145/epi.2018.mar.21. Disponível em: <https://arxiv.org/pdf/1804.05365>. Acesso em: 15 mar. 2021.

ORDUÑA-MALEA, E. *et al.* *La revolución Google Scholar: destapando la caja de Pandora académica*. Madrid: Unión de Editoriales Universitarias Españolas, 2016.

ORTEGA, J. L. *Academic search engines: a quantitative outlook*. [S. l.]: Elsevier, 2014.

PEREIRA, F. A. *Mapeamento do uso do Google Scholar em estudos bibliométricos ou cientométricos: análise bibliométrica e revisão sistemática*. 2022. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2022. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27151/tde-12042023-125534/publico/FabianaAndradePereiraCorrigida.pdf>. Acesso em: 15 mar. 2024.

SANTOS, G. C. *Fontes de indexação para periódicos científicos: um guia para bibliotecários e editores*. Campinas: E-Color, 2011. Disponível em: <http://eprints.rclis.org/16898/1/fontesISBN9788563058188.pdf>. Acesso em: 15 dez. 2020.

SANTOS, S. M. dos *et al.* The relationship between the language of scientific publication and its impact in the field of public and collective health. *Journal of Scientometric Research*, [s. l.], v. 10, n. 1, p. S78-S87, Jan./Abr. 2021. DOI 10.5530/jscires.10.1s.24. Disponível em: <https://jscires.org/article/419>. Acesso em: 14 jul. 2021.

SAUVAYRE, R. Types of Errors Hiding in Google Scholar Data. *Journal of Medical Internet Research*, [s. l.], v. 24, n. 5, p. 1-19, 2022. Disponível em: <https://www.jmir.org/2022/5/e28354/PDF>. Acesso em: 28 maio 2022.

SILVA, D. D.; ALMEIDA, C. C. de; GRÁCIO, M. C. C. Associação do Fator de Impacto e do Índice h para a avaliação de periódicos científicos: uma aplicação no campo da Ciência da Informação. *Em Questão*, Porto Alegre, v. 24, p. 132-151, 2018. DOI 10.19132/1808-5245240.132-151. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/86489/52329>. Acesso em: 25 maio 2021.

SILVA, D. D.; GRACIO, M. C. C. Índice h de Hirsch: análise comparativa entre as bases de dados Scopus, Web of Science e Google Acadêmico. *Em Questão*, Porto Alegre, v. 23, p. 196-212, 2017. Edição especial 5º EBBC. DOI 10.19132/1808-5245230.196-212. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/68010/40074>. Acesso em: 5 mar. 2021.

SINGH, V. K. *et al.* The journal coverage of Web of Science, Scopus and Dimensions: a comparative analysis. *Scientometrics*, [Budapest], v. 126, p. 5113-5142, 2021. DOI 10.1007/s11192-021-03948-5. Disponível em:

<https://link.springer.com/article/10.1007/s11192-021-03948-5>. Acesso em: 15 mar. 2021.

SOCIEDADE BRASILEIRA DE MEDICINA TROPICAL. *Fator de Impacto e Índice H para avaliar publicações e autores*. Brasília, DF, 9 de maio de 2020. Disponível em: <https://sbmt.org.br/fator-de-impacto-e-indice-h-para-avaliar-publicacoes-e-autores/>. Acesso em: 25 maio 2021.

THELWALL, M. Does Microsoft Academic find early citations? *Scientometrics*, [Budapest], v. 114, p. 325-334, 2018b. DOI 10.1007/s11192-017-2558-9. Disponível em: <https://link.springer.com/article/10.1007/s11192-017-2558-9>. Acesso em: 15 fev. 2021.

THELWALL, M. Microsoft Academic: A multidisciplinary comparison of citation counts with Scopus and Mendeley for 29 journals. *Journal of Informetrics*, [s. l.], v. 11, n. 4, p. 1201-1212, 2017. DOI 10.1016/j.joi.2017.10.006. Disponível em: <https://arxiv.org/pdf/1711.08767>. Acesso em: 15 fev. 2021.

THELWALL, M. Microsoft Academic automatic document searches: Accuracy for journal articles and suitability for citation analysis. *Journal of Informetrics*, [s. l.], v. 12, n. 1, p. 1-9, 2018a. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1751157717303346>. Acesso em: 15 fev. 2021.

THOMAZ, P. G.; ASSAD, R. S.; MOREIRA, L. F. P. Uso do Fator de impacto e do índice h para avaliar pesquisadores e publicações. *Arquivos Brasileiros de Cardiologia*, São Paulo, v. 96, n. 2, p. 90-93, 2011. Disponível em: <https://www.scielo.br/j/abc/a/kp6RzbScSJt5snkS7XQvsqy/?format=pdf&lang=pt>. Acesso em: 2 ago. 2021.

TORRES-SALINAS, D.; JIMÉNEZ-CONTRERAS, E. Hacia las unidades de bibliometría en las universidades: modelo y funciones. *Revista Española De Documentación Científica*, [Madrid], v. 35, n. 3, p. 469-480, 2012. DOI 10.3989/redc.2012.3.959. Disponível em: <https://redc.revistas.csic.es/index.php/redc/article/view/753/834>. Acesso em: 15 fev. 2021.

TORRES-SALINAS, D.; RUIZ-PÉREZ, R.; DELGADO-LOPEZ-CÓZAR, E. Google Scholar como herramienta para la evaluación científica. *El Profesional de la Información*, [Barcelona], v. 18, n. 5, p. 501-510, sept./oct. 2009. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2009.sep.03/21564>. Acesso em: 15 jun. 2020.

VARGAS, R. A.; VANZ, S. A. de S.; STUMPF, I. R. C. The role of national journals on the rise in Brazilian agricultural science publications in Web of Science. *Journal of Scientometric Research*, [s. l.], v. 3, n. 1, p. 28-36, Jan./Apr. 2014. Disponível em: <https://jscires.org/article/64>. Acesso em: 25 ago. 2022.

VERA-BACETA, M.-A.; THELWALL, M.; KOUSHA, K. Web of Science and Scopus language coverage. *Scientometrics*, [Budapest], v. 121, n. 3, p. 1803-1813, 2019. DOI 10.1007/s11192-019-03264-z. Disponível em: <https://link.springer.com/article/10.1007/s11192-019-03264-z>. Acesso em: 15 fev. 2021.

WALTERS, W. H. Google Scholar coverage of a multidisciplinary field. *Information Processing & Management*, New York, v. 43, n. 4, p. 1121-1132, jul. 2007. DOI 10.1016/j.ipm.2006.08.006. Disponível em: https://www.researchgate.net/publication/222411552_Google_Scholar_coverage_of_a_multidisciplinary_field/link/6599ad740bb2c7472b376634/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19. Acesso em: 12 mar. 2021.

CAPÍTULO 6

Avaliação por pares em periódicos científicos: características e importância para evitar periódicos predatórios

Edna da Silva Angelo

Marlene Oliveira

Introdução

Sendo configurado como um dos meios de comunicação mais eficientes para a transferência da informação e o desenvolvimento econômico e social de uma sociedade, o periódico científico, desde os seus primórdios, desenvolveu-se como resposta funcional aos problemas de assimetria de comunicação. Isso ocorreu por causa de dois fatores: o tempo – determinado pela velocidade de distribuição – e o espaço – considerando a área de distribuição.

Com base no modo com que o ciclo da publicação ocorre, o cientista, o pesquisador ou o docente são avaliados, profissionalmente, pela produção de seus artigos. Publicar é prova de atividade científica e acadêmica, o que pode garantir uma boa avaliação do pesquisador pelas universidades e agências de fomento. Assim, o periódico científico promove o reconhecimento da atividade de pesquisa (Biojone, 2003) e representa o principal meio para conferir prestígio aos autores (Stumpf, 1998).

Esse instrumento utilizado na comunicação científica manifesta-se como um mecanismo para grupos acadêmicos compartilharem seus estudos, todavia, diferente de outras formas de divulgação, a arbitragem permite que as informações fornecidas tenham um grau relativamente alto de confiabilidade. Isso tornou-se de extrema importância em face dos desafios atuais de excesso e incerteza relacionados à veracidade do conteúdo disponibilizado na *web*.

Nesse sentido, este capítulo apresenta reflexões sobre as características dos periódicos, a existência dos periódicos predatórios, a importância da equipe editorial e da avaliação por pares e a composição do processo de avaliação para garantia de qualidade das pesquisas publicadas e da credibilidade do periódico. Trata-se de uma temática vinculada à área da ciência da informação, em específico o campo dos estudos direcionados ao processo da comunicação científica.

Periódicos científicos

A palavra “publicação” é de origem latina do vocábulo *publicatione* e o seu significado relaciona-se ao ato ou efeito de publicar; já a palavra “periódico” vem do latim *periodus*, que significa período ou espaço de tempo. Assim, publicações periódicas são as informações disseminadas de tempos em tempos, atendendo a uma frequência regular de fascículos ou números, sob um mesmo título, dentro de uma área específica do conhecimento e/ou de amplitude global. Tais publicações devem atender às normas e padrões internacionais que permitam a visibilidade e o reconhecimento do que é publicado (Fachin, 2006; Meadows, 1999).

O periódico é um instrumento de informação que revela pensamentos, ações, desejos, manifestações e ideais. Apresenta quatro principais razões de ser: informar, interpretar, orientar e entreter. Além disso, o periódico ajuda a conhecer os atores; traz um panorama da história sobre os acontecimentos políticos, sociais e culturais de um país; e manifesta, em suas linhas, os pensamentos, as ideias circulantes e os acontecimentos cotidianos de uma época (Landa Landa, 2006).

No entendimento de Rodrigues e Marinho (2009), a publicação no formato periódico desempenha um papel eficaz em ajudar a comunicar novas pesquisas à comunidade científica e à própria sociedade como um todo. Nessas publicações, segundo Rodrigues e Marinho (2009), são divulgados os resultados dos estudos mais atuais sobre determinado tema, tornando-se o meio mais ágil de atualização da informação científica e tecnológica. Sendo assim, a transferência do conhecimento por meio dos periódicos é,

sem dúvida, um elemento-chave para o avanço de uma determinada ciência ou disciplina em particular (Restrepo-Valencia *et al.*, 2015).

Caracterizados como meios de comunicação científica, os periódicos adotam um sistema de comunicação científica com elemento estruturador vital da comunidade científica (Mueller, 2017). Miranda e Pereira (1996) delinearam duas vertentes que caracterizaram a importância dos periódicos como veículos formais de comunicação da ciência: 1) comunicação do conhecimento e 2) comunicação entre os pares. Para os autores, é a partir desses dois aspectos que os periódicos devem ser observados no tocante à visibilidade e à credibilidade, pois proporcionam divulgação com maior rapidez e abrangência.

Por intermédio desses canais, a pesquisa é formalizada e se torna conhecimento público (Mabe, 2001; Meadows, 1999; Miranda; Pereira, 1996). Sobre isso, Meadows (1999) menciona que a comunicação se situa no próprio coração da ciência. É tão vital quanto à própria pesquisa, pois a esta não cabe reivindicar legitimidade se não houver sido analisada e aceita pelos pares. Isso exige, essencialmente, que seja comunicada. Ademais, o apoio às atividades científicas é dispendioso, e os recursos que lhes são alocados serão desperdiçados a menos que os resultados das pesquisas sejam mostrados aos públicos pertinentes. Qualquer que seja o ângulo pelo qual a examinemos, a comunicação efetiva constitui parte essencial do processo de investigação científica.

Em 1665, como pegadas da ciência moderna, surgiram os primórdios do atual periódico científico. As duas primeiras publicações originaram-se nesse ano e com apenas 60 dias de diferença, de um lado e de outro do Canal da Mancha. Estas foram o *Journal des Sçavans* (França) e as *Philosophical Transactions* (Inglaterra), mais tarde renomeadas para *Journal des Savants* e *Philosophical Transactions of the Royal Society*, respectivamente (Leite, 2008; Spinak; Packer, 2015).

Para Stumpf (1996), desde que começaram a ser publicados, os periódicos passaram a desempenhar importante papel no processo de comunicação da ciência. A autora destaca que eles surgiram como uma evolução do sistema particular e privado de comunicação que era feito por meio de cartas

compartilhadas entre os investigadores e de atas ou memórias das reuniões científicas.

Esse movimento provocou um aspecto organizacional distintivo e vital na revolução científica, cristalizando um novo conjunto de normas, incentivos e estruturas organizacionais que reforçaram os compromissos dos pesquisadores científicos com a divulgação rápida de novos conhecimentos. Com base no novo conceito e nas características das publicações periódicas, cresceram miríades de estruturas parcialmente institucionalizadas e parcialmente comercializadas. Tais estruturas desenvolvidas constituíram o contexto cultural, político e básico no qual os conhecimentos acadêmicos se fundamentaram (David, 2004).

À luz dos estudos de Jange e Kademani (1999), Kronick (1976) e Mueller (2000), os periódicos científicos ajudam a manter o padrão de qualidade da ciência, pois os manuscritos estão sujeitos ao processo de revisão pelos pares, a fim de ser garantida a qualidade diante do número crescente de fraudes ou erros. Essa revisão confere confiabilidade ao material a ser publicado, de acordo com esses autores. Na concepção de Mueller (2000), a confiabilidade é uma das características mais importantes da ciência, pois possibilita, assim, diferenciar o conhecimento popular do científico.

Como uma publicação formal validada, Mabe (2001), Meadows (1999) e Stumpf (1996) afirmam que o periódico mantém o registro científico da investigação e dos seus resultados, atuando como instrumento que preserva o conhecimento científico. Nesse sentido, Stumpf (1998) define os periódicos como o arquivo da ciência, enquanto Castedo e Gruszynski (2011) afirmam que, mesmo em formato digital, esses canais mantêm a capacidade de preservação e que essa característica confere aos periódicos especial importância na cadeia de difusão e de construção do conhecimento científico de cada campo ou área do conhecimento.

Considerando que os periódicos publicam resultados de pesquisa e que esses resultados serão preservados para constituição do conhecimento científico, surge outra característica peculiar e muito importante: a avaliação. Nesse sentido, Spudeit, Werlang e Presser (2012) defendem que a incumbência da avaliação é coletar evidências significativas e fornecer à equipe editorial informações necessárias e bases para a tomada de decisão em

relação à melhoria do processo editorial, oferecendo a esta as condições de acompanhar e monitorar os próprios resultados dos estudos, com vistas a alcançar níveis de qualidade compatíveis com os padrões da área na qual o periódico está inserido.

A avaliação é um processo que serve de filtro, que busca eliminar pesquisas consideradas de baixa qualidade ou que não sejam apresentadas de forma adequada. Mesmo tendo o seu valor, o processo é alvo de críticas, sobretudo relacionadas ao grande tempo despendido. É muito provável que iniciativas de mudança ou melhoria nesse procedimento continuem a surgir, aproveitando as vantagens oferecidas pelas tecnologias digitais, conforme a velocidade com que o conhecimento cresce ou com a urgência de disponibilizar esse conhecimento ao público (Moreno Ceja; Cortés Vera; Zumaya Leal, 2012).

Mesmo que haja o desejo de melhorar ou de modificar o processo avaliativo, a princípio, não está contemplado o estabelecimento de uma prática que possa substituir a avaliação por pares, visto que tal mudança não seria facilmente aceita pelas comunidades científicas das diversas áreas da ciência. Outro fator relevante é estabelecer a propriedade intelectual, ou seja, ao publicar seus estudos em um periódico, o autor registra, de modo formal, a sua autoria, requerendo para si a descoberta científica (Mueller, 2000; Rodrigues; Oliveira, 2012; Roosendaal; Geurts, 1997).

Tradicionalmente, o desempenho científico dos pesquisadores é medido de duas maneiras: a) quantos artigos o pesquisador publicou, que fornece uma medida da produtividade absoluta; e b) quantas vezes esses estudos foram citados pelos outros artigos, ou seja, constaram da sua lista bibliográfica, que avalia o impacto e a importância do artigo em seu campo científico (Ohira; Sombrio; Prado, 2000). O ato de publicar artigos é exigido pelos pares como prova definitiva de efetiva atividade em pesquisa científica (Miranda; Pereira, 1996).

Com efeito, entende-se que a publicação dos resultados de pesquisa em padrões autorizados e referendados por um periódico científico constitui não somente uma característica da ciência e um direito do pesquisador, mas um dever de se fazer ciência. Esse comportamento é esperado pelos pares e pelos empregadores ou instituições a que o pesquisador está vinculado.

Periódicos predatórios

Atentos à disposição de alguns autores a pagar para publicar seus manuscritos de modo rápido, surgiram periódicos que não são comprometidos com a ética tampouco com a qualidade da publicação científica. Assiste-se ao crescimento de um mercado comercial altamente lucrativo, cujo autor paga para ter seu artigo publicado, sem uma criteriosa revisão por pares. Essas publicações, com modelos de negócios questionáveis, são denominadas como periódicos predatórios (Prado; Kraenkel; Coutinho, 2017).

Na visão de Logullo (2016), a principal missão desses periódicos é ganhar dinheiro e não contribuir para o avanço da ciência, pois disponibilizam textos quase sem nenhum rigor, desde que os autores paguem. A autora citada destaca que os periódicos predatórios exploram o modelo de acesso aberto Article Processing Charge (APC), ou a taxa de processamento de artigo, cobrando a taxa sem fornecer um processo de revisão por pares de qualidade. A partir disso, Butler (2013) relata que foi identificado, desde 2008, um fluxo crescente de mensagens de novos periódicos solicitando artigos. Entretanto, a maioria dos *e-mails* continha vários erros gramaticais e, ao navegar nos *sites* dos periódicos, verificou-se que seus editores não eram exatamente o que afirmavam. Os nomes muitas vezes soavam grandiosos – palavras do tipo “mundo”, “global” e “internacional” eram comuns –, mas alguns *sites* pareciam amadores ou forneciam poucas informações sobre a organização por trás deles.

O relatório da InterAcademy Partnership (2022), que teve o objetivo de identificar intervenções viáveis e efetivas para conter e ajudar a combater o aumento preocupante de periódicos predatórios e oferecer recomendações às comunidades envolvidas, descreve que o impacto da pandemia do coronavírus (covid-19) pode ser uma causa a mais para preocupação. Isso porque, conforme o relatório, as pesquisas de má qualidade e atores inescrupulosos capitalizam a confusão e a urgência causadas pela pandemia, considerando que surgiram, no contexto pandêmico, novas formas de gestão, de comunicação e de uso das pesquisas.

A maioria das publicações predatórias não apresenta padronização e nem transparência adequadas. Assim, passou-se a divulgar listas de alerta

quanto à existência de publicações duvidosas. Um exemplo dessa iniciativa é o trabalho de Jeffrey Beall, biblioteconomista da Universidade de Colorado – em Denver, nos Estados Unidos –, ao montar uma página na internet que lista centenas de periódicos de várias áreas e que foram considerados como predatórios. Essa iniciativa pioneira foi denominada Lista de Beall ou *Beall's List*, e gera conhecimento para os pesquisadores, sobretudo os menos experientes (Logullo, 2016).

O campo da metapesquisa tem se dedicado a investigar mais a fundo o fenômeno dos periódicos predatórios e a Lista de Beall já não é mais considerada atualizada, nem única. Esse crescimento está relacionado ao aumento, também, do número de periódicos duvidosos (Cabells Scholarly Analytics, 2022). Assim, surgiram outros critérios para caracterizar um periódico como predatório, propostos por equipes diferentes, muitas vezes até discordantes. Assim, o consenso existe em classificar periódicos científicos predatórios como aqueles que cobram taxas para a publicação, não praticam revisão por pares (*peer review*), têm procedimentos obscuros ou de baixa qualidade de revisão – portanto, publicam “qualquer coisa” – e não são indexados nas bases de dados utilizadas para a pesquisa (Logullo, 2016).

Observa-se que essas listas podem ser contraproducentes, pois é difícil produzir classificações desse tipo de modo adequado e reunir uma relação completa de publicações suspeitas. Fato é que pesquisadores deveriam receber treinamento para distinguir publicações confiáveis (Fundação de Amparo à Pesquisa do Estado de São Paulo, 2017). Prado, Kraenkel e Coutinho (2017) indicam que a eficiência para detectar periódicos potencialmente predatórios varia muito entre campos de pesquisa e recomendam que todas as áreas do conhecimento explicitem critérios específicos para identificar potenciais publicações predatórias.

Conforme relatado pela InterAcademy Partnership (2022), existe uma grande confusão e uma série de mal-entendidos sobre o que constitui periódicos predatórios. Com o aumento, a diversificação e a sofisticação das práticas de pesquisa, a instituição mencionada destaca que é cada vez mais difícil identificar, rastrear, monitorar e distinguir com segurança os periódicos fraudulentos, inferiores ou de baixa qualidade. A InterAcademy Partnership (2022) apresenta um espectro de comportamentos que evita a

classificação simplista de periódicos bons ou ruins, seguros ou em observação, e válidos ou inválidos. Infere-se que esses comportamentos não são perfeitos, mas objetivam garantir a transparência, ao invés de ocasionar o julgamento por valor. Isso porque o valor de certos traços pode ser diferente em contextos geográficos distintos, sobretudo pelos traços serem dinâmicos e intensificados pela sua rápida evolução e pela complexidade cada vez maior (InterAcademy Partnership, 2022).

A coordenação da área de educação na Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), no âmbito da nota sobre o Qualis Periódicos da área da educação, menciona que promove avaliação cuidadosa de periódicos chamados predatórios. A coordenação considera como predatórios aqueles que assediam os pesquisadores, cobrando taxas de submissão/publicação elevadas e garantindo prazo muito curto para a aprovação e publicação do artigo, além de não indicarem, com clareza, se há rigorosa avaliação cega entre os pares (Brasil, 2021).

Um fator que justifica o crescimento dos periódicos predatórios é a valorização da quantidade em detrimento à qualidade. Ribeiro (2014) mencionou essa tendência em entrevista concedida ao periódico *Nature*, em 2014, no artigo intitulado “Architects of South American Science”. Conforme explica a pesquisadora, no Brasil, as duas das maiores barreiras quanto à melhoria da pesquisa do país são a avaliação de desempenho e as recompensas. A valorização da quantidade em vez da qualidade está tão arraigada na cultura científica brasileira que é apelidada de numerologia. Em vez de reunir um conjunto completo de experimentos em uma história coerente, os cientistas ganham mais reconhecimento no sistema ao dividir ou fragmentar o estudo em várias produções (Ribeiro, 2014).

O mesmo apontamento é mencionado por Santos-d’Amorim (2021), ao relatar sobre a equação, composta, de um lado, por pesquisadores vivendo sob a máxima “*publique ou pereça*” e, do outro, a percepção da indústria da informação científica como um vantajoso nicho de mercado, o que faz elevar a quantidade de editoras predatórias. A autora conclui que as ofertas direcionadas por *e-mail*, prometendo rápida publicação, têm levado à mercantilização do artigo científico como uma moeda de troca e, junto com ela, manifesta-se, também, a pressão ao pesquisador.

A avaliação por pares é a área menos apoiada e menos documentada do processo de pesquisa. A ausência de clareza e de limpidez no processo, originalmente desenvolvido para minimizar a parcialidade do sistema por meio de confidencialidade, permite que as práticas predatórias sigam despercebidas e incontestadas. Além desses fatores, a escassez de reconhecimento profissional e de treinamento para avaliação por pares gera a desmotivação para que alguém sirva como avaliador, o que, por conseguinte, faz com que incentivos para pegar atalhos e reduzir o rigor tornem-se ainda mais atraentes à promessa de serviços predatórios (InterAcademy Partnership, 2022).

Para mitigar essas ocorrências, o relatório da InterAcademy Partnership (2022) aborda algumas ações direcionadas a cinco diferentes agentes, a saber:

- *governos e financiadores de pesquisa*: que esses agentes encomendem mais pesquisas sobre a avaliação por pares, a fim de ajudar a promover padrões, além de entender como funciona o processo avaliativo e como ele pode evoluir no futuro;
- *instituições de educação*: que elas garantam o reconhecimento profissional das funções editoriais e de avaliação por pares, em igualdade com a publicação de artigo, ou outra forma de produção escrita, e tornem esses aspectos parte dos critérios de análise de pesquisa, promoção e reconhecimento. Ademais, que sejam atribuídos pontos de bonificação a serem utilizados em taxas de acesso aberto;
- *financiadores de pesquisa*: que promovam a publicação em periódicos com transparência e/ou com processos de avaliação aberta por pares e vinculem apoio financeiro para a publicação, nesses periódicos;
- *editoras*: que façam da avaliação transparente por pares uma norma – copublicação do relatório de avaliação –, com flexibilidade na divulgação de nomes e do nível de sigilo do processo de avaliação. Além disso, que desenvolvam programas de treinamento e códigos de conduta de boas práticas para a avaliação por pares;
- *Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) e outras organizações intergovernamentais*: que aumentem a conscientização sobre os benefícios da avaliação transparente por pares, a ser realizada de modo aberto, anônimo ou híbrido.

A partir do crescimento dos periódicos predatórios e do perigo que eles representam para a qualidade da ciência, confirma-se a importância de se manter os periódicos científicos, com uma estrutura sólida que possibilite o fortalecimento do seu funcionamento e da sua gestão (Squazzoni, 2020). Como destacado por Silveira e Silva (2020), é por meio da valorização dos periódicos científicos, munidos de estruturas adequadas e fortalecidas, que o processo de avaliação da ciência atingirá os seus objetivos, tendo em vista qualificar o que é produzido pela ciência.

Equipe editorial científica

Os periódicos têm como objetivo o pleno acesso à informação de qualidade. Para o cumprimento desse propósito, faz-se necessária a organização da equipe editorial, caracterizada como o conjunto de colaboradores que exercem atividades distintas e integradas, possibilitando a constituição do fluxo editorial. Como reforçado por Guirro (2010), o trabalho dessa equipe constitui a própria essência do periódico. Portanto, a publicação depende de inúmeros colaboradores de diversas áreas profissionais. O processo começa com os pesquisadores submetendo os manuscritos para avaliação, passando pela estrutura administrativa, a qual viabiliza o andamento do processo desde o recebimento dos manuscritos até a publicação final, tendo a participação, nessas etapas, dos editores especialistas, pareceristas, editores-chefes e o pessoal da preparação de texto e produção gráfica.

Sobre a constituição da equipe editorial, Trzesniak (2009) salienta que tal processo é extremamente importante e, ao mesmo tempo, complexo, sobretudo por depender das diferentes circunstâncias a que cada instituição mantenedora do periódico está inserida, como disponibilidade de recursos, profissionais qualificados, dentre outras condições. O autor menciona também, como fator relevante, o conhecimento sobre as diversas nomenclaturas atribuídas aos membros das equipes, dependendo da atividade realizada.

Mesmo ciente dessa afirmação do autor, a realidade empírica de muitos periódicos, sobretudo os brasileiros, tem evidenciado a falta de padronização no uso de termos para designar processos, funções e, principalmente, nomes atribuídos aos grupos que participam das diferentes atividades. No

que tange aos nomes atribuídos aos grupos ou indivíduos que se envolvem com o processo avaliativo, apresenta-se o Quadro 1, que exemplifica a composição da estrutura da equipe editorial de um periódico, estrutura essa que visa contribuir para a melhoria do fluxo editorial.

Quadro 1 – Estrutura da equipe editorial de um periódico

1	comissão editorial
2	editor-chefe
3	editor associado
4	conselho editorial
5	parecerista <i>ad hoc</i>
6	editor honorário
7	equipe técnica

Fonte: elaborado pelas autoras (2023).

O conjunto terminológico padronizado exposto no Quadro 1 foi definido após uma pesquisa de doutorado, considerando os termos mais utilizados pelos periódicos de acesso aberto do Brasil e da Espanha, resultado inicial alcançado por meio de pesquisa documental realizada nos *websites* de 2.628 periódicos brasileiros e 1.619 espanhóis. O modelo proposto foi avaliado como adequado para suprir as necessidades informacionais dos editores científicos, possibilitando, assim, uma indexação eficiente do periódico.

A contribuição prática do produto alcançado com a tese – modelo terminológico – foi evitar que as designações sejam estabelecidas por conveniência, sem critérios seletivos. Os sete grupos estabelecidos são detalhados, na sequência, considerando, sobretudo, as informações coletadas durante a pesquisa documental, no estudo de doutorado ora citado, além de discussões teórico-conceituais oriundas da literatura publicada.

A *comissão editorial*, entendida como um colegiado autônomo, é responsável pelas questões administrativas e pela política editorial, a qual deve conter as seguintes informações: missão, escopo e foco, público-alvo, identidade visual, processo editorial e periodicidade. A comissão editorial é constituída por um grupo de pessoas que elaboram as diretrizes, estabelecendo o perfil político-editorial do periódico.

Essa equipe não tem envolvimento com o conteúdo de qualquer artigo em particular, mas com a coleção como um todo, estabelecendo as normas gerais que a norteiam. Os membros da comissão são consultados sobre o tema ou sobre temas correspondentes a cada número publicado. Além disso, eles avaliam o resultado de cada edição, atuam em bloco – fazem reuniões periodicamente, de modo presencial ou *on-line* –, tomam decisões coletivas e, eventualmente, apresentam propostas.

Uma das funções da equipe é fornecer orientação intelectual ao periódico e, se for o caso, a outras atividades administrativas como aprovar a gestão econômica e fiscalizar a atuação de seus órgãos executivos. Outrossim, pode auxiliar na indicação de avaliadores e nas demais decisões relativas aos aspectos científicos do processo editorial do periódico. De qualquer modo, escolher o editor-chefe é uma das principais atribuições dessa comissão.

A incumbência da comissão editorial é discutir e desenvolver regulamentos próprios para assuntos variados. Dentre os temas, destacam-se: normas e formatação de artigos, critérios de avaliação e aceite de textos para publicação, aceite de novos grupos de pesquisa para participar do comitê, aprovação do orçamento, credenciamento de revisores e pareceristas, definição de política de circulação e assinaturas, entre outros.

Ter uma comissão editorial é muito saudável, pois divide os encargos do editor-chefe com outros colaboradores sobre as decisões quanto ao rumo do periódico. Carrillo Fuentes (2020) menciona que a equipe editorial constitui o órgão colegiado por meio do qual são discutidos os manuscritos recebidos e se decide enviá-los ou não para a fase de parecer. Essa é a fase mais delicada, pois será definido se o estudo é científico e cumpre os requisitos do periódico. Em caso positivo, segundo a autora citada, o manuscrito será avaliado, na maioria das vezes, por, pelo menos, dois especialistas no assunto.

A comissão editorial é composta por uma equipe de profissionais da instituição e de outras universidades, familiarizados com os processos editoriais e comprometidos com a tarefa de alcançar um aumento progressivo da qualidade da publicação. Os membros devem possuir experiência em pesquisa, como autor e como avaliador, principalmente nas áreas do conhecimento relacionadas ao foco do periódico.

O número de membros é definido no regulamento do periódico que, em alguns casos, é elaborado pela instituição responsável. Geralmente, a comissão é composta por membros internos e externos da editora. Esses membros podem ser nacionais e internacionais e podem viver na mesma cidade ou fora dela. No caso de pesquisadores que moram em outro país ou que estão fora do país, a tecnologia pode ser utilizada para garantir a continuidade dos trabalhos (Carrillo Fuentes, 2020).

A comissão em apreço precisa estar permanentemente ativa no cumprimento das definições gerais da política editorial do periódico e dos procedimentos editoriais previstos, sobretudo no que se refere à garantia de consistência e de qualidade das publicações. Edições especiais e outras contribuições científicas deverão ser discutidas e definidas por essa instância.

O *editor-chefe* é o encarregado pela gestão do periódico e desempenha um papel central no processo de controle ético, de avaliação e de publicação da pesquisa científica. É o agente responsável por tudo que é veiculado no periódico, desde a gestão editorial até o desempenho final com as edições publicadas, conduzindo as políticas editoriais decididas pela comissão editorial. Ademais, cabe a ele a revisão preliminar das políticas, o encaminhamento dos manuscritos aos avaliadores, além da representação institucional e a iniciativa de propor os conteúdos fundamentais de cada número.

Em correspondência com Carrillo Fuentes (2020), uma das funções do editor-chefe é observar, com rigor, as regras de acompanhamento e de avaliação de um artigo científico e propor a organização dele e como ela pode ser otimizada para elevar a publicação a um nível acadêmico superior. Esse colaborador fará uma revisão técnica do manuscrito recebido, confirmará se atende aos padrões editoriais do periódico e se segue as diretrizes da linguagem científica. Também realizará, segundo o autor citado, uma primeira revisão do aparato crítico, devendo consultar o autor para completar as informações que faltam, fiscalizar a adequada qualidade do material gráfico, entre outros detalhes.

O acompanhamento respeitoso do cumprimento das normas editoriais agrega seriedade aos periódicos. O editor-chefe conhece a correção do estilo editorial – como um texto é formado gráfica e tipograficamente – e faz as revisões necessárias para aprovar a publicação. Esse agente do processo

editorial conhece todo o fluxo, desde o momento em que um autor submete seu manuscrito ao periódico até o momento da disponibilização do estudo para consulta (Carrillo Fuentes, 2020).

Assim como Carrillo Fuentes (2020), Silva, Moreiro-González e Mueller (2016) destacam as atribuições e a importância do editor-chefe, mencionando, sobremaneira, que a importância e a responsabilidade desse agente estão não somente na preservação dos interesses e das políticas editoriais, mas também por ser ele um agente de vigilância da qualidade do conhecimento científico, de modo geral. Assim, o editor-chefe precisa ter uma postura ativa no desenvolver das atividades e responderá aos autores eventuais dúvidas sobre o processo de submissão, em quaisquer de suas etapas. Todas as submissões são, a princípio, consideradas pelo editor-chefe, que verifica o atendimento aos requisitos formais da submissão, sua adequação à política editorial e seu potencial de publicação. Somente os estudos considerados como relevantes para a comunidade e, em particular, para os leitores do periódico prosseguem para as demais etapas de avaliação.

A tarefa substancial do editor-chefe de um periódico soma-se aos seus múltiplos compromissos como pesquisador e professor, pois nenhuma dessas tarefas lhe é eximida, o que implica dividir-se para cumprir todas as responsabilidades. Em algumas instituições, tem sido discutida a necessidade de esse agente se concentrar, apenas, no trabalho do periódico, mas isso implica alguma controvérsia, devido à importância do seu envolvimento com a atividade de pesquisa, pois isso lhe confere experiência e, com efeito, permite-lhe cumprir com excelência as obrigações editoriais (Carrillo Fuentes, 2020).

Esse trabalho está estreitamente vinculado ao do parecerista, pois o que se espera desse último é que avalie com justiça, equilíbrio e clareza o conteúdo do manuscrito, emitindo comentários e pareceres com as devidas justificativas. Assim, em conformidade com as reflexões de Silva, Moreiro-González e Mueller (2016), a análise dos manuscritos, antes do envio para a revisão por pares, é uma atividade que compete ao editor-chefe, mas pode também ser realizada por editores associados, membros do conselho editorial ou órgão equivalente, se o periódico possuir. Com isso, esses autores mencionam que a avaliação inicial realizada pelo editor-chefe afere o

interesse editorial, a qualidade geral do manuscrito e o grau de contribuição do estudo.

Como suporte à atividade avaliativa, o editor-chefe, após avaliação inicial, identifica revisores especialistas para realizar a avaliação profunda dos manuscritos. Como a revisão de artigos deve ser realizada por pares ou especialistas no tema de que trata o estudo analisado, identificar o pesquisador mais apropriado para atuar como parecerista reflete na qualidade do periódico. Isso requer do editor amplo domínio de sua área do conhecimento para que possa encontrar, entre as especialidades e os temas que compõem o escopo do periódico, os pareceristas que dominam os temas abordados nos estudos a serem avaliados. Portanto, a literatura entende que o papel do editor – como também dos pareceristas – é determinante para a configuração de uma política editorial consistente e coerente com aquilo que propõe o periódico, como também para a busca da qualidade da produção científica (Silva; Moreiro-González; Mueller, 2016).

A atuação do editor-chefe carece de uma formação estruturada e especializada para enriquecer o trabalho que realiza. Considerando que há carência de uma formação especializada, caberá a ele recorrer a uma formação continuada, que contemple conhecimentos sobre edição, ferramentas técnicas e gestão de *software*. Ademais, precisa conhecer os programas antiplágio, as aplicações para *design* editorial e a marcação eletrônica, além de ser treinado para fazer marcação em arquivos XML (Carrillo Fuentes, 2020).

Outras necessidades do editor-chefe, de acordo com Carrillo Fuentes (2020), são buscar conhecimento sobre as tendências e necessidades de indexação de periódicos e padrões de qualidade, apropriar-se do uso de números que identificam textos, imagens, gráficos, dentre outros objetos, e realizar o acompanhamento para custear a gestão e registrar os novos fascículos. Carrillo Fuentes (2020) ressalta, ainda, que o editor precisa ser treinado para conhecer gerenciadores bibliográficos, termos bibliométricos e usos de bibliometria e cienciometria, como também conhecer as fontes de consulta – por exemplo, os repositórios, os diretórios, entre outras –, trabalho esse que poderá ser realizado em parceria com a biblioteca da instituição.

Para treinar um novo editor-chefe é demandado tempo e esforço para que o profissional consiga conhecer e dominar o processo de gestão

editorial. O processo de aprendizagem é permanente e o trabalho de um periódico, especialmente científico e arbitrado, é dinâmico, imparável e muito rigoroso. A importância da preparação desse agente para realizar as tarefas que lhe competem em um periódico está relacionada com as dificuldades que enfrentam, pois o que se espera é que editores mais preparados tenham menos dificuldades (Silva; Moreiro-González; Mueller, 2016). Essa preparação, relatam os autores, envolve tanto aspectos técnicos quanto científicos e requer investimento constante na busca de novos conhecimentos científicos e editoriais referentes, principalmente, à gestão editorial, ao uso de tecnologias, à ciétiometria, entre outros.

Silva, Moreiro-González e Mueller (2016) buscaram detectar o nível de dificuldade enfrentado por editores na realização da revisão por pares em periódicos científicos que tenham sido avaliados como de qualidade, bem como o nível de preparação desses editores na realização das tarefas diárias. A avaliação das dificuldades dos editores considerou, inicialmente, oito tarefas principais, a saber:

- realização da avaliação dos manuscritos antes do envio para a revisão pelos pares;
- identificação de revisores especialistas e preparados para realizar a avaliação dos manuscritos;
- promoção do envolvimento do conselho editorial na seleção dos manuscritos antes de serem submetidos à revisão pelos pares;
- determinação do grau de originalidade do manuscrito apresentado;
- identificação de conflitos de interesse na escolha dos revisores;
- resolução de conflitos entre revisores – quando há recomendações divergentes acerca do manuscrito;
- orientação dos revisores quanto ao tipo de análise que interessa ao periódico;
- decisão sobre o destino de manuscritos que ficaram sem revisores.

A identificação dessas tarefas, segundo Silva, Moreiro-González e Mueller (2016), demonstra que o trabalho editorial é complexo, permeado por responsabilidade e envolvimento por parte, sobretudo, do editor-chefe. Os autores finalizam afirmando que, embora sejam atividades frequentes na gestão de periódicos, não correspondem à totalidade de tarefas que realizam os editores.

O editor-chefe pode delegar atribuições específicas do processo editorial ao editor associado, que tem como atribuição colaborar na gestão do fluxo editorial. Esse agente não é imprescindível no âmbito das boas práticas da publicação científica, mas muito pode contribuir quando há um número alto de submissões admitidas no processo editorial (Associação Nacional de Pós-Graduação e Pesquisa em Administração, 2010).

O *editor associado* tem a missão de tratar dos assuntos burocráticos em relação ao periódico. Também pode ser o encarregado de receber o texto e encaminhá-lo aos avaliadores especialistas que julgar apropriados. Cumpre a função de assessorar o editor-chefe, em especial, nos assuntos relativos à direção executiva.

Dentre as funções atribuídas a esse agente do processo editorial, citam-se: auxiliar a direção do periódico em todas as suas funções, especialmente no acompanhamento dos manuscritos recebidos – recepção, avaliação, aceitação e revisão final – e atuar na definição do conteúdo e do estilo do periódico – verificação e aplicação das regras de submissão dos originais, concepção, criação e orientação das diferentes seções do periódico. Além disso, o editor associado acompanha o fluxo dos manuscritos e garante o cumprimento das normas e preceitos éticos estabelecidos.

O *conselho editorial* é o órgão encarregado pela avaliação dos estudos recebidos e pela pertinência deles no que se refere ao conteúdo e à qualidade das contribuições. Os originais serão submetidos à apreciação de membros do conselho editorial do periódico, de acordo com a especificidade do tema e do domínio de conhecimento do membro do conselho.

Esse conselho é composto por pesquisadores que apresentam relevantes contribuições ao campo ou tema de pesquisa. Na respectiva área de especialidade, essa equipe atua como consultora do editor-chefe, no que diz respeito à interpretação da política editorial, instituída pela comissão editorial.

Esses membros não se reúnem para discutir algum assunto ou tomar alguma decisão, pois são pesquisadores que trabalham sozinhos e interagem, separadamente, com o editor. Eles se envolvem com o conteúdo dos artigos, preocupando-se com a uniformidade, a continuidade, a qualidade e o rigor científico do que é publicado (Trzesniak, 2009).

Esta é a principal função da revisão por pares: fornecer um parâmetro sobre a qualidade do manuscrito para que o editor tenha mais segurança na decisão final acerca da publicação do manuscrito. Entretanto, se não foi possível submeter o manuscrito à revisão por pares, o editor-chefe deve decidir sobre o destino desse material (Silva; Moreiro-González; Mueller, 2016).

Em síntese, o aconselhamento e a supervisão da qualidade científica dos conteúdos do periódico constituem o papel principal do conselho editorial. Nesse ínterim, essa equipe é composta por professores universitários e pesquisadores de reconhecido prestígio e por membros de instituições educacionais ou governamentais nacionais e internacionais e a sua missão é colaborar na elaboração da estratégia científica e na avaliação periódica dos resultados, bem como apoiar e orientar o processo, de maneira conceitual e científica.

O *conselho editorial* reserva-se ao direito de indicar especialistas com evidente competência acadêmica nas áreas que não são integrantes da equipe editorial, como o parecerista *ad hoc*. A equipe dispõe de plena autoridade para decidir sobre a conveniência ou não da publicação, podendo inclusive reapresentá-la aos autores, com sugestões para que sejam feitas alterações necessárias no texto e/ou para que os textos sejam adaptados às normas editoriais do periódico.

O grupo de *pareceristas ad hoc* é composto por pessoa(s) convidada(s) a analisar os artigos. Esse grupo não faz parte do corpo editorial científico e é consultado para avaliar determinado artigo submetido à revista. Não possui, no entanto, nenhum vínculo permanente com o periódico e encerra suas participações ao término da avaliação do artigo que lhe foi designado. São pesquisadores nacionais e estrangeiros – preferencialmente com abrangência em todos os continentes – e reconhecidos nas áreas dos periódicos que colaboram, sobretudo no que tange ao assunto dos manuscritos que avaliam (Trzesniak, 2009).

Como abordado por Silva, Moreiro-González e Mueller (2016), a identificação de avaliadores pelos editores dos periódicos requer amplo domínio de sua área do conhecimento para que possam encontrar, dentre as várias especialidades e temas, o avaliador adequado para cada manuscrito. Os autores concluíram que, entre os editores de periódicos considerados de qualidade e em variadas áreas do conhecimento no Brasil, na Espanha e no México, a tarefa de localizar avaliadores foi considerada difícil por eles.

O recrutamento desses pareceristas se dará a partir de seu conhecimento e sua formação acadêmica, que deve contemplar uma das subáreas abrangidas pelo periódico. São membros convidados, responsáveis pela avaliação dos artigos, acadêmicos competentes nas diversas áreas da publicação e classificados em pares acadêmicos por temas específicos. A seleção deles é realizada com critérios rigorosos de afinidade e reconhecimento de sua trajetória investigativa. Para essa seleção, a equipe editorial conta com uma sólida base de dados em nível internacional e é auxiliada por especialistas, sobretudo da área da comunicação científica, que ajudarão a encontrar possíveis pareceristas, incluindo os três principais idiomas: espanhol, inglês e português.

Os *editores honorários* são nomeados pela comissão editorial e, na ausência desse grupo, pelo editor-chefe. Tais membros farão parte do referido conselho honorário enquanto não renunciarem, explicitamente, ou forem revogados. Professores universitários e pesquisadores de reconhecido prestígio serão nomeados membros do conselho honorário. A presença não é obrigatória, pois não atuam no fluxo editorial, nem na gestão do periódico, e não avaliam artigos. São personalidades referenciadas por motivo honorário ou para agregar prestígio.

A *equipe técnica* auxilia nas tarefas da gestão administrativa e financeira. Os manuscritos aceitos são encaminhados para edição, quando deverão ser inseridas todas as informações autorais do estudo pelos autores e ocorrerá a revisão do resumo, tanto na língua nativa como estrangeira, por equipe própria. Posteriormente, os manuscritos seguirão para a lista do número do periódico. No Brasil e na Espanha, o termo mais utilizado para essa categoria é editor assistente ou secretário.

Em geral, as funções do editor assistente ou secretário são efetivar as contribuições e as propostas recebidas dos membros da comissão editorial e, juntamente com o editor-chefe, responsabilizar-se pela localização, gestão e manutenção das bases de indexação e das listas de distribuição adequadas ao periódico. Outros encargos recorrentes são: normalização bibliográfica, revisão gramatical, tradução, edição de texto, diagramação, edição de *layout*, suporte do sistema, emissão das certidões necessárias – atestando a comprovação dos fatos acreditados –, redação da ata da reunião anual, envio dos originais aos avaliadores e preparação de relatórios estatísticos anuais sobre o processo editorial.

Em resumo, nota-se que o papel de cada membro da equipe é fundamental para aprimorar a qualidade da publicação. As atividades desempenhadas são essenciais para operar o processo editorial. Assim, é importante atentar que, durante as atividades voltadas à avaliação, o periódico deve prezar, sobretudo, pelo diálogo estabelecido entre aquele que expõe suas ideias e o avaliador. Para tanto, em geral, os pareceres não têm apenas o caráter de rejeição ou aprovação das submissões, mas serão propostos com o fim de provocar um diálogo construtivo que compreenderá análises quanto ao tema explorado no estudo.

Processo de avaliação por pares

A avaliação por pares, ou revisão por pares (*peer review*), estabelecida por uma equipe editorial científica, é uma etapa anterior à publicação dos resultados das pesquisas sendo imprescindível ao sistema de comunicação científica. Nessa lógica, esta resulta em um ganho substancial porque confere o compartilhamento de saberes de excelência e dele se vale para apresentar conteúdos informacionais que sustentam a geração do conhecimento.

Sendo assim, por meio da análise à estrutura editorial dos periódicos brasileiros e espanhóis, foi possível desenvolver um fluxograma das atividades editoriais de avaliação de artigos que, em geral, passam por um processo composto de três fases: pré-avaliação, avaliação por partes e revisões finais.

A primeira fase compreende a pré-avaliação ou análise prévia. Ao receber a submissão de um manuscrito, o editor-chefe cumpre uma primeira

fase de apreciação, que envolve a avaliação da correspondência da proposta de publicação com o foco e o escopo do periódico e dos aspectos da formatação. Havendo alguma irregularidade, o manuscrito é devolvido aos autores; não havendo, o artigo será encaminhado para apreciação do conselho editorial ou *consejo asesor*, que poderá fazer uso de pareceristas *ad hoc* ou *evaluador* de reconhecida competência na área do conhecimento.

A segunda fase é a avaliação pelos pares, conduzida por pareceristas a convite do editor-chefe. Os avaliadores serão selecionados com base em suas áreas específicas de conhecimento, em correspondência com as temáticas tratadas nos artigos. Com base nas recomendações e pareceres exarados nas avaliações, o conselho editorial ou *consejo asesor* apresentará uma proposta de decisão ao editor-chefe, que poderá:

- *aprovar o texto para publicação*: o estudo é aceito integralmente para publicação em um dos próximos números do periódico, seguindo o critério cronológico de conclusão do processo de análise;
- *solicitar aos autores alterações no texto*: as modificações deverão ser realizadas pelos autores, que receberão o parecer com as referidas recomendações;
- *rejeitar o texto para publicação*: o manuscrito é rejeitado, com a devida justificativa atribuída pelo editor-chefe, que terá como referência a análise feita pelos revisores.

É atribuição do editor-chefe acatar ou não as recomendações dos pareceristas e encaminhar eventuais pedidos de reformulação aos autores. A decisão final sobre a publicação de um manuscrito sempre será do editor-chefe e da comissão editorial, que farão uma avaliação do texto original, das sugestões indicadas pelos consultores e das modificações encaminhadas pelo autor.

Por fim, a terceira fase é destinada às revisões finais, à publicação e à divulgação. O texto aprovado seguirá para a atividade de edição – preparação final para publicação – e será submetido às seguintes ações: revisão das normas bibliográficas – nessa atividade, a equipe editorial reserva-se o direito de proceder à revisão gramatical dos textos e fazer correções, desde

que não alterem o conteúdo; revisão dos descritores do artigo e dos metadados; e composição do texto diagramado, conforme o padrão gráfico da editora.

Cumpridas as três fases do processo de avaliação, os textos serão encaminhados aos autores para aprovação da revisão. Não havendo manifestação, o texto será considerado aprovado e será publicado.

Considerações finais

As reflexões propostas neste capítulo acerca do amplo e completo processo de avaliação em periódicos científicos levam a reforçar que o periódico tem a missão de promover o enriquecimento do debate científico, além de servir como meio de divulgação a um público mais amplo. Com isso, pauta-se pela avaliação criteriosa para atingir tal objetivo, garantindo a maior neutralidade possível dos pareceres e a seleção de textos de qualidade sobre a temática. Os manuscritos submetidos serão admitidos para publicação após avaliação por pares, desde que atendam ao escopo geral do periódico e às normas para publicação, mesmo no caso de edições que, porventura, privilegiem algum tema em especial.

Nas primeiras décadas do século XXI, é possível perceber que os periódicos, ao longo do tempo, ganharam destaque: por um lado, pelos avanços tecnológicos e, por outro, pelas demandas de produtividade na avaliação da atividade científica. O valor do artigo, principal gênero textual veiculado pelos periódicos, como aborda Vilchis, Rivera-López e Martínez-Domínguez (2020), está relacionado, principalmente, às mudanças ocorridas nos meios de circulação do conhecimento, nos modos de compreender a internacionalização da pesquisa, nos altos níveis de profissionalização e na modernidade dos sistemas que administram e avaliam a ciência, tendo em vista o financiamento dela e o avanço na carreira dos pesquisadores.

O papel desempenhado pelos periódicos científicos e o debate em torno desses meios de comunicação da ciência evidenciam que eles constituem um mecanismo para grupos acadêmicos compartilharem seus estudos, mas, diferentemente de outras formas de divulgação, o processo de arbitragem permite que as informações publicadas nos artigos tenham um

grau relativamente alto de qualidade e confiabilidade. Assim, os artigos são submetidos a um criterioso processo de avaliação, por meio de revisão por pares. Essa etapa é concebida com o propósito de garantir que estudos de qualidade, fidedignos, inéditos, sem vieses e que contribuam para o processo da ciência sejam publicados.

A avaliação por pares é considerada como a estratégia mais efetiva para garantir a qualidade, confiabilidade, integridade e consistência da literatura acadêmica. Desse modo, essa etapa busca, a partir do conhecimento dos especialistas convidados (pareceristas ou revisores), avaliar a importância e a adequação da informação científica que está sendo veiculada, conferindo, portanto, confiabilidade à informação. Muito mais do que um simples processo de avaliação e de atribuição de um “selo” de qualidade aos manuscritos submetidos, o processo de avaliação visa, também, ao aperfeiçoamento do texto e de sua adequação às diretrizes dos periódicos.

Entende-se que o processo de revisão por pares transparente e eficaz é algo satisfatório para a ciência, visto que tende a diminuir as inconsistências nos resultados das pesquisas, como também os casos de má conduta, tanto de autores quanto dos próprios periódicos, eliminando ou reduzindo, assim, a existência de periódicos de má qualidade ou com interesses não científicos. Além disso, evidencia-se que o desenvolvimento da ciência e a necessidade de sua comunicação por meio do artigo científico exigem profissionalização, reconhecimento por suas instituições e especialização dos editores.

Referências

ARCHITECTS of South American science. *Nature*, [s. l.], v. 510, p. 209-212, June 2014. Disponível em: <https://www.nature.com/articles/510209a>. Acesso em: 19 mar. 2023.

ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO. *Boas práticas da publicação científica: um manual para autores, revisores, editores e integrantes de corpos editoriais*. Rio de Janeiro: ANPAD, 2010. Disponível em: https://inpeau.paginas.ufsc.br/files/2013/01/manual_boas_praticas_anpad.pdf. Acesso em: 19 mar. 2023.

BIOJONE, M. R. *Os periódicos científicos na comunicação da ciência*. São Paulo: EdUC: FAPESP, 2003.

BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. *Nota sobre o Qualis Periódicos: área de Educação*. Brasília, DF, jan. 2021.

BUTLER, D. The darkside of publishing: The explosion in open-access publishing has fuelled the rise of questionable operators. *Nature*, [s. l.], v. 495, p. 433-435, Mar. 2013. Disponível em: <https://www.nature.com/articles/495433a>. Acesso em: 19 mar. 2023.

CABELLS SCHOLARLY ANALYTICS. *Predatory reports*. [S. l.], 2022.

CARRILLO FUENTES, Rosalba. Como se conforma una revista científica y arbitrada? In: JIMÉNEZ YÁÑEZ, C. E. (coord.) *Revistas académicas em ciências sociais y humanidades en México: realidades, experiencias y expectativas*. Baja California: Universidad Autónoma de Baja California, 2020. p. 236-247. Disponível em: <https://www.colson.edu.mx/publicaciones/files/15034.pdf>. Acesso em: 19 mar. 2023.

CASTEDO, R. da S.; GRUSZYNSKI, A. A produção editorial de revistas científicas online: uma análise de publicações brasileiras da área da Comunicação. *Em Questão*, Porto Alegre, v. 17, n. 1, p. 271-287, jan./jun. 2011. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/16697/12482>. Acesso em: 19 mar. 2023.

DAVID, P. A. Understanding the emergence of ‘open Science’ institutions: functionalist economics in historical context. *Industrial and Corporate Change*, [Oxford], v. 13, n. 4, p. 571-589, Aug. 2004.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. Nova lista de periódicos predatórios. *Pesquisa FAPESP*, São Paulo, ed. 256, jun. 2017. Disponível em: <https://revistapesquisa.fapesp.br/nova-lista-de-periodicos-predatorios/>. Acesso em: 19 mar. 2023.

GUIRRO, R. O trabalho em equipe do processo editorial. *Fisioterapia em Pesquisa*, São Paulo, v. 17, n. 2, p. 101, jul./dez. 2010. DOI 10.1590/S1809-29502010000200001. Disponível em: <https://www.scielo.br/j/fp/a/hk4S5mLjBm4pp3SBgDmTv7G/?format=pdf&lang=pt>. Acesso em: 7 abr. 2023.

INTERACADEMY PARTNERSHIP. *Combating Predatory Academic Journals and Conferences*. [S. l.]: IAP, 2022. Disponível em: <https://www.>

interacademies.org/sites/default/files/2022-03/1.%20Full%20report%20-%20English%20FINAL.pdf. Acesso em: 19 mar. 2023.

JANGE, S.; KADEMANI, B. S. Metamorphosis of the scientific journal: past, present and future. *Malaysian Journal of Library & Information Science*, [s. l.], v. 4, n. 1, p. 61-69, July 1999. Disponível em: <http://eprints.rclis.org/4638/1/article.pdf>. Acesso em: 19 mar. 2023.

KRONICK, D. *A history of scientific and technical periodicals*. 2th nd. Metuchen: Scarecrow Press, 1976.

LANDA LANDA, M. G. Publicaciones antiguas mexicanas (1805-1950). *Biblioteca Universitária*, México, v. 9, n. 1, p. 09-15, enero/jun. 2006. Disponível em: <http://www.redalyc.org/pdf/285/28590103.pdf>. Acesso em: 19 mar. 2023.

LEITE, M. P. F. R. O que é um periódico científico de qualidade? *Femina*, [s. l.], v. 36, n. 12, dez. 2008.

LOGULLO, P. As revistas open access predatórias. *Palavra Impressa*, [s. l.], 5 ago. 2016. Disponível em: <https://www.palavraimpressa.com.br/2016/08/05/as-revistas-open-access-predatorias/>. Acesso em: 19 mar. 2023.

MABE, M. A. Digital dilemmas: electronic challenges for the scientific journal publisher. *Aslib Proceedings*, Bradford, v. 53, n. 3, p. 85-92, Mar. 2001.

MEADOWS, A. J. *A comunicação científica*. Brasília, DF: Briquet de Lemos, 1999.

MIRANDA, D. B. de; PEREIRA, M. de N. F. O periódico científico como veículo de comunicação: uma revisão de literatura. *Ciência da Informação*, Brasília, DF, v. 25, n. 3, p. 375-382, set./dez. 1996. Disponível em: <https://revista.ibict.br/ciinf/article/view/636/640>. Acesso em: 19 mar. 2023.

MORENO CEJA, F.; CÓRTES VERA, J. de J.; ZUMAYA LEAL, M. del R. Usos, limitaciones y prospectiva de la evaluación por pares. *Revista Interamericana de Bibliotecología*, Medellín, v. 35, n. 2, p. 201-210, 2012. Disponível em: . Acesso em: 19 mar. 2023.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. *Ciência da Informação*, Brasília, DF, v. 46, n. 3, p. 49-50, maio/ago. 2017. DOI 10.1590/S0100-19652006000200004. Disponível em: <https://www.scielo.br/j/ci/a/>

nGD3MkKfNxtjnnWshf3YVjP/?format=pdf&lang=pt. Acesso em: 19 mar. 2023.

MUELLER, S. P. M. O periódico científico. In: CAMPELLO, B. S.; CENDÓN, B. V.; KREMER, J. (org.). *Fontes de informação para pesquisadores e profissionais*. Belo Horizonte: Ed. UFMG, 2000. p. 73-96.

OHIRA, M. L. B.; SOMBRIO, M. L. L. N.; PRADO, N. S. Periódicos brasileiros especializados em Biblioteconomia e Ciência da Informação: evolução. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 10, n. 10, p. 01-17, out. 2000. Disponível em: <https://www.redalyc.org/articulo.oa?id=14701004>. Acesso em: 19 mar. 2023.

PRADO, P. I.; KRAENKEL, R. A.; COUTINHO, R. M. *Preda Qualis*: periódicos potencialmente predatórios no QUALIS-CAPES. Brasília, DF, 2017. Disponível em: <https://predaqualis.netlify.app/resultados/>. Acesso em: 19 mar. 2023.

RESTREPO-VALENCIA, L. *et al.* Análisis de laproducción científica de la Revista CES Odontología em los últimos 10 años. *Revista CES Odontología*, Medellín, v. 28, n. 2, p. 119-31, 2015. Disponível em: <https://revistas.ces.edu.co/index.php/odontologia/article/view/3682/2490>. Acesso em: 19 mar. 2023.

RIBEIRO, S. Reward quality not quality. *Nature*, [s. l.], v. 510, p. 211, June 2014. Disponível em: <https://www.nature.com/articles/510209a>. Acesso em: 19 mar. 2023.

RODRIGUES, J. G.; MARINHO, S. M. O. X. A trajetória do periódico científico na Fundação Oswaldo Cruz: perspectivas da Biblioteca de Ciências Biomédicas. *História, Ciências, Saúde - Manguinhos*, Rio de Janeiro, v. 16, n. 2, p. 523-532, abr./jun. 2009. Disponível em: <https://www.scielo.br/pdf/hcsm/v16n2/16.pdf>. Acesso em: 19 mar. 2023.

RODRIGUES, R. S.; OLIVEIRA, A. B. de. Periódicos científicos na America Latina: títulos em acesso aberto indexados no ISI e SCOPUS. *Perspectiva em Ciência da Informação*, Belo Horizonte, v. 17, n. 4, p. 77-99, out./dez. 2012. Disponível em: <http://www.scielo.br/pdf/pci/v17n4/06.pdf>. Acesso em: 19 mar. 2023.

ROOSENDAAL, H. E.; GEURTS, P. A. T. M. *Forces and functions in scientific communication: an analysis of the irinter play*. [S. l.: s. n.], 1997. Disponível em: https://www.researchgate.net/publication/237407169_Forces_and_functions_in_scientific_communication_an_analysis_of_their_interplay/link/00b7d5283ec13cdcb8000000/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InNpZ251cCIsInBhZ2UiOiJwdWJsaWNhdGlvbiJ9fQ. Acesso em: 19 mar. 2023.

SANTOS-D'AMORIM, K. A comunicação científica em movimento: das origens aos debates atuais. *Brazilian Journal of Information Science*, Marília, v. 15, p. 01-20, 2021. Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/11468/7041>. Acesso em: 19 mar. 2023.

SILVA, C. N. N. da; MOREIRO-GONZÁLEZ, J. A.; MUELLER, S. P. M. A revisão por pares a partir da percepção dos editores: um estudo comparativo em revistas brasileiras, espanholas e mexicanas. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 14, n. 1, p. 126-143, jan./jun. 2016. DOI 10.20396/rdbci.v14i1.8640579. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8640579/pdf>. Acesso em: 19 mar. 2023.

SILVEIRA, L. da; SILVA, F. C. C. da (org.). *Gestão editorial de periódicos científicos: tendências e boas práticas*. Florianópolis: EdUFSC, 2020. Disponível em: <https://repositorio.ufc.br/handle/riufc/52892>. Acesso em: 19 mar. 2023.

SPINAK, E.; PACKER, A. L. 350 anos de publicação científica: desde o “Journal des Sçavans” e “Philosophical Transactions” até o SciELO. *SciELO em Perspectiva*, São Paulo, 5 mar. 2015. Disponível em: <https://blog.scielo.org/blog/2015/03/05/350-anos-de-publicacao-cientifica-desde-o-journal-des-savans-e-philosophical-transactions-ate-o-scielo/>. Acesso em: 19 mar. 2023.

SPUDEIT, D.; WERLANG, E.; PRESSER, N. H. Indicadores de gestão do fluxo editorial dos periódicos científicos: uma reflexão teórico-metodológica. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 17, p. 102-117, 2012. Número especial. DOI 10.5007/1518-2924.2012v17nesp2p102. Disponível em: <https://cip.brapci.inf.br//download/32526>. Acesso em: 19 mar. 2023.

SQUAZZONI, F. Avaliação por pares não é apenas controle de qualidade, é parte integrante da infraestrutura social da pesquisa. *SciELO em Perspectiva*, São Paulo, 15 jan. 2020. Disponível em: <https://blog.scielo.org/blog/2020/01/15/avaliacao-por-pares-nao-e-apenas-controle-de-qualidade/#.YyvKIHbMLIU>. Acesso em: 19 mar. 2023.

STUMPF, I. R. C. Passado e futuro das revistas científicas. *Ciência da Informação*, Brasília, DF, v. 25, n. 3, p. 383-386, set./dez. 1996. DOI 10.18225/ci.inf.v25i3.637. Disponível em: <https://revista.ibict.br/ciinf/article/view/637/641>. Acesso em: 19 mar. 2023.

STUMPF, I. R. C. Reflexões sobre as revistas brasileiras. *Intexto*, Porto Alegre, v. 1, n. 3, p. 1-10, jan./jun. 1998. Disponível em: <https://seer.ufrgs.br/intexto/article/view/3369/3953>. Acesso em: 19 mar. 2023.

TRZESNIAK, P. A estrutura editorial de um periódico científico. In: SABADINI, A. A. Z. P.; SAMPAIO, M. I. C.; KOLLER, S. H. (org.). *Publicar em Psicologia: um enfoque para a revista científica*. São Paulo: Associação Brasileira de Editores Científicos de Psicologia; Instituto de Psicologia da Universidade de São Paulo, 2009. p. 87-102.

VILCHIS, I. L.; RIVERA-LÓPEZ, B.; MARTÍNEZ-DOMÍNGUEZ, N. D. Más allá del paper: géneros discursivos en revistas académicas de humanidades. In: JIMÉNEZ YÁÑEZ, C. (coord.). *Revistas académicas en ciencias sociales y humanidades en México: realidades, experiencias y expectativas*. Baja California: Universidad Autónoma de Baja California, 2020. p. 81-102. Disponível em: <https://www.colson.edu.mx/publicaciones/files/15034.pdf>. Acesso em: 19 mar. 2023.

CAPÍTULO 7

Repositórios institucionais de produção científica: agregando valor com a implementação de indicadores relacionais complexos

Rubenildo Oliveira da Costa

Lúcia Maria Sebastiana Verônica Costa Ramos

Introdução

Os repositórios subdividem-se em institucionais, governamentais, disciplinares e agregadores, conforme classificação do Repositório Global de Repositórios de Acesso Aberto (OpenDOAR, 2023). De modo mais específico, os Repositórios Institucionais (RIs) de universidades disponibilizam em acesso aberto as seguintes coleções produzidas pela sua comunidade – docentes/pesquisadores/discentes/funcionários: produção científica publicada, *preprints* – conteúdo científico não revisado por pares –, dados – dados gerados nas pesquisas, tais como relatórios, arquivos, planilhas etc. –, teses e dissertações, material didático, memória institucional etc.

Os RIs, por meio de bibliotecas universitárias, coletam, registram, certificam, armazenam, preservam e promovem a visibilidade e o uso dos mais variados documentos inseridos nas coleções citadas anteriormente. Desse modo, podem ser considerados como a fonte de informação mais abrangente da produção científica de uma instituição, superando as maiores bases de dados mundiais, as quais têm o foco na coleção de artigos científicos.

Além da importância do controle e visibilidade, eles podem fornecer também indicadores bibliométricos a partir dos registros bibliográficos, tais como: quantidade de publicações por autor, departamento, unidade, assunto etc., que permitem uma análise da produção científica institucional

mais abrangente – variada tipologia de publicações – e *mais específica* – possibilidade de restrição por unidade, departamento, programas de pós-graduação, grupos de pesquisa etc.

Contudo, mesmo considerando todo o potencial e tempo de proliferação dos RIs nos últimos 20 anos, constatam-se preocupações ainda com adesão, visibilidade e uso de tal recurso, e poucos estudos sobre indicadores na literatura internacional dos últimos cinco anos.

Em vista disso, pretende-se aqui contribuir com a discussão acerca dos porquês do “baixo” desempenho – visibilidade e uso – dos RIs, observada na literatura recente, além da necessidade de implementação e dos usos de indicadores bibliométricos mais complexos.

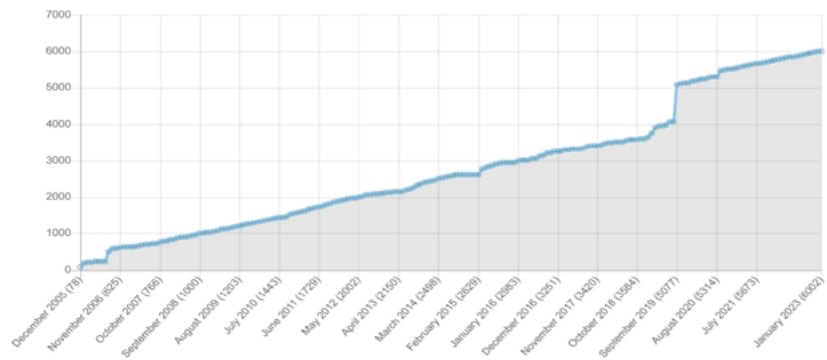
Adesão, visibilidade e uso de RIs

A proposição da implementação de indicadores relacionais complexos requer o entendimento da situação atual dos RIs, notadamente as implicações sobre sua adesão, visibilidade e uso e relações existentes com o movimento de acesso aberto, evolução tecnológica, surgimento de novas tipologias de documentos, além da importância do papel das bibliotecas universitárias no controle, coleta, registro, certificação e preservação da produção científica, acadêmica e cultural da instituição.

Adesão

O avanço na implantação de RIs de acesso aberto no mundo nos últimos 20 anos é observado na literatura, notadamente relatado em publicações de estudos de casos e no *site* do OpenDOAR, conforme a Figura 1 a seguir:

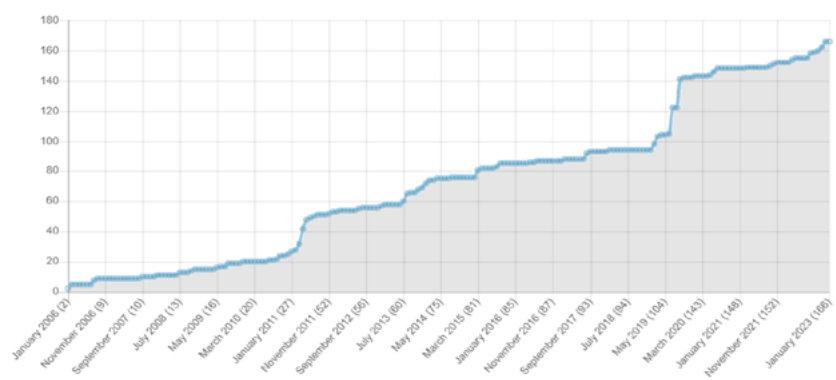
Figura 1 – Crescimento de RIs de acesso aberto



Fonte: OpenDOAR (2023).

Do mesmo modo, foi observado um crescimento no Brasil, conforme a Figura 2 a seguir:

Figura 2 – Crescimento de RIs de acesso aberto no Brasil



Fonte: OpenDOAR (2023).

Tal crescimento pode estar relacionado, a princípio, a três fatores: ampliação do movimento de acesso livre, avanço tecnológico e inclusão de novas tipologias/coleções de documentos.

Considerado o principal fator de crescimento de RIs, o Movimento de Acesso Livre, segundo Luther (2018), iniciou-se em 2002 com a Budapest

Open Access Initiative, na qual instituições e acadêmicos assumiram o compromisso de autoarquivar suas publicações e estimular a criação de revistas de acesso aberto, o que resultou na criação da Declaração de Budapest. Tal movimento contrapõe-se ao modelo tradicional de publicação científica baseado em altos custos de assinaturas de periódicos de editoras comerciais, e considera a premissa de que o conhecimento científico financiado com recursos públicos deve ser de livre acesso à sociedade, agilizando, assim, a pesquisa, a descoberta e a inovação na ciência.

No Brasil, a Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), seguindo as premissas do movimento de acesso livre e a exemplo de outras agências de fomento internacionais, divulgou em 2019 uma portaria que dispõe sobre a recomendação da publicação científica em acesso aberto resultantes de pesquisas realizadas com recursos da agência, e que deverá ser armazenada em um RI, caso não seja publicada em periódico de acesso aberto (Fundação de Amparo à Pesquisa do Estado de São Paulo, 2019).

Por outro lado, as editoras comerciais, atentas às iniciativas internacionais de acesso aberto, adaptaram-se e implementaram o modelo híbrido de publicação nos seus periódicos de acesso fechado, nos quais os autores pagam para publicar em acesso aberto. Ou seja, os periódicos de acesso fechado também publicam artigos de acesso aberto – via híbrida. Quando o autor não dispuser de condições de pagar para publicar em acesso aberto em um periódico fechado, a editora comercial pode autorizá-lo a armazenar a versão aprovada do artigo em um repositório de acesso aberto, os chamados *postprint*, ou mesmo a versão não aprovada, os *preprints* – via verde. Desse modo, os RIs se tornam peça-chave para as vias alternativas do acesso aberto (García Peñalvo, 2017).

Sendo assim, observa-se que tal movimento provocou um aumento crescente e significativo no número de publicações científicas em acesso aberto nos últimos anos. Como exemplo, pode-se citar o fato de que, pela primeira vez, o número de publicações em acesso aberto superou o de acesso fechado em 2019, segundo a plataforma Dimensions (2023).

Outro fator importante para o aumento no número de RIs no mundo é o avanço tecnológico, que começou também em 2002, quando o Massachusetts Institute of Technology (MIT) criou o *software* de código

aberto DSpace para criação de repositórios de acesso livre (Luther, 2018). Por ser um *software* livre de código aberto, o DSpace viabilizou a popularização de RIs. Atualmente, é responsável por 39% dos repositórios no mundo, segundo o OpenDOAR.

No Brasil, o DSpace foi customizado e disponibilizado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) em 2004, que também oferece suporte, assim como outras redes de bibliotecas nacionais, para a construção de um RI utilizando tal *software*.

O último fator importante é o aumento na tipologia de conteúdo armazenado nos RIs. Ou seja, além dos documentos tradicionalmente armazenados, tais como artigos, livros, capítulos de livros, teses e dissertações e resumos de congressos, surgiram outras demandas para as instituições, tais como o armazenamento dos seguintes documentos: objetos educacionais, *preprints/postprints*, dados de pesquisas científicas, entre outros.

A variedade de tipologia de documentos e integração entre os RIs tem o potencial de alterar a comunicação científica atualmente dominada por publicadores comerciais, segundo o grupo de trabalho da Confederação de Repositórios de Acesso Aberto (2017). No entanto, observa-se que a implementação dessas novas coleções – *preprints/postprints*, dados de pesquisa, objetos educacionais etc. – ainda está nos seus estágios iniciais, pois demanda aperfeiçoamento tecnológico, alteração cultural dos produtores de conteúdos e consolidação conceitual e técnica.

Diante do exposto, acredita-se que tais fatores – ampliação do movimento de acesso livre, avanço tecnológico e inclusão de novas tipologias de documentos – contribuíram e ainda contribuem para a adesão aos RIs, uma vez que as instituições sentem cada vez mais a necessidade de criar e/ou aperfeiçoar seus repositórios, seja pela exigência de órgãos superiores de financiamento de pesquisa/ensino ou pelo próprio potencial dessa importante fonte de informação.

Visibilidade e uso

A literatura internacional recente sobre os RIs apresenta benefícios e problemas relacionados à visibilidade e uso da produção científica das instituições.

Dentre os benefícios mais comuns, destaca-se o fato de, ao reunir a produção científica em uma única fonte de informação, evitar-se sua dispersão em várias bases de dados, facilitando assim a busca da informação, além de permitir mais visibilidade e uso. Essa vantagem se amplia no caso de documentos não indexados, tais como: livros, capítulo de livros, resumos de congresso etc.

Ademais, a disponibilização da produção científica institucional em outras fontes de informação por meio de interoperabilidade – entre RIs – é também um dos fatores que influencia a sua visibilidade e uso (Moreira *et al.*, 2017; Shearer, 2017). Um dos exemplos disso é a disponibilização da produção no Google – principal buscador na internet.

As ferramentas que buscam o Formato Portátil de Documento (PDF) em vários RIs no mundo desempenham um papel importante na recuperação da informação (Azadbakht; Schultz, 2020; Dhakal, 2019), tais como as extensões Unpaywall¹, EndNoteClick² e Mendeley³, permitindo a obtenção de um documento por meio de apenas um clique.

Contudo, apesar dessas principais vantagens, observam-se problemas relacionados à visibilidade e uso, que dificultam atingir toda a potencialidade dos RIs. Sendo assim, o principal fator, que influencia no armazenamento e na disponibilização dos documentos nos RIs, é a restrição imposta pelos direitos autorais de editoras comerciais (Shearer, 2017; Wu, 2015). Por isso, os RIs possuem mais registros bibliográficos do que documentos digitais em si. Com isso, é possível constatar com uma simples busca que o resultado será de inúmeros registros bibliográficos sem os respectivos documentos anexos. Nesse sentido, a coleção de teses e dissertações é a campeã de acesso e *downloads* dentre as coleções, por disponibilizarem o texto integral. Portanto, entende-se que quanto mais documentos armazenados, maior será o seu uso.

Existe também o problema da coleta e registro de um número crescente de produção científica por um número cada vez menor de bibliotecários, o

1 Disponível em: <https://unpaywall.org/products/extension>.

2 Disponível em: <https://click.endnote.com/>.

3 Disponível em: <https://www.mendeley.com/download-reference-manager>.

que influencia na cobertura do conteúdo. Nesse ponto, é importante ressaltar a importância das bibliotecas universitárias na coleta, no registro e na certificação da produção científica, diante da insuficiência do autoarquivamento por parte dos pesquisadores (Luther, 2018). Alguns motivos dessa insuficiência do autoarquivamento, segundo Tay (2017), reside no desconhecimento da importância dos RIs e dos direitos autorais por parte dos pesquisadores, além da falta de motivação/incentivos da instituição.

As atividades de coleta e registro da produção são minuciosas e morosas. Coletar a produção exige uma busca exaustiva em várias bases de dados, além da criação de alertas e de uma seleção rigorosa sobre se realmente aquela produção pertence a um autor da instituição. Já o registro demanda tempo e atenção na identificação e registro dos elementos bibliográficos, tais como autor, título, ano etc.

Contudo, existem ferramentas que podem minimizar as atividades de coleta e registro da produção, tais como os identificadores digitais únicos, como, por exemplo, o Open Researcher and Contributor ID (ORCID) do autor, que pode auxiliar na coleta; e o Digital Object Identifier System (DOI) do documento, que permite a importação dos elementos bibliográficos dos documentos para o registro nos RIs. Entretanto, estas não são ferramentas abrangentes, enquanto a primeira carece de atualização por parte do autor, a segunda é limitada a apenas documentos digitais.

Por último, os RIs perdem para os Repositórios Temáticos em termos de funcionalidades, inovação e aceitação por parte dos pesquisadores (Tay, 2017). Nesse sentido, a inteligência artificial é considerada um diferencial entre eles (Meleiro, 2018).

Diante do exposto, é possível perceber a relação de causa e efeito entre os problemas apontados anteriormente na disponibilização de documentos – principalmente de direitos autorais e avanços tecnológicos – e na baixa visibilidade e no uso de RIs. Sendo assim, torna-se crucial agregar valor aos RIs para aumentar a sua importância na instituição, diminuindo a dependência do movimento de acesso aberto para seu desenvolvimento e evolução.

Trata-se de mudança de foco de disponibilização de documentos em PDF para a oferta de informação estratégica, para tomada de decisão.

Ou seja, ao considerar que os RIs possuem mais registros bibliográficos do que documentos armazenados, sugere-se aqui aumentar a visibilidade e o uso de tal recurso, focando na ampliação do número de indicadores bibliométricos relacionais de coocorrência de dados, o que permite uma visão mais ampla das atividades acadêmicas dos autores a partir de representações gráficas dinâmicas em redes.

Da necessidade de implementação de novos indicadores relacionais complexos por meio da integração entre sistemas internos e externos

Uma alternativa viável para aumentar a importância, a visibilidade e o uso de RIs, em contrapartida à disponibilização de PDFs com problemas decorrentes, é a implementação de indicadores relacionais complexos para análise e/ou avaliação da produção científica institucional.

Indicadores relacionais complexos estabelecem relações entre documentos, autores, instituições etc., ao invés da soma de métricas individuais. Eles permitem uma avaliação mais completa da produção científica e podem ser úteis na identificação de padrões de colaboração, influência e transferência de conhecimento; podem fornecer, por exemplo, uma visão da estrutura e evolução de pesquisas interdisciplinares, conforme observado no estudo de Youngblood e Lahti (2018).

Em estudo recente, Reis, Spinola e Amaral (2017) realizaram um levantamento dos principais indicadores disponíveis nos RIs brasileiros, por serem considerados importantes fontes de informação para a elaboração de estudos bibliométricos. Os autores concluíram que a maioria dos indicadores é de produção e altimétricos, e poucos de citação. Os indicadores de produção se referem à contagem de publicação por autor, ano, instituição etc., enquanto os de citação, ao número de citações recebidas por publicação e indicadores decorrentes, tais como o fator e índice de impacto e o índice h. Os indicadores altimétricos se referem à contagem de visualizações e *downloads* por documentos.

No mesmo estudo, Reis, Spinola e Amaral (2017) revelaram ainda a incipiência de representações gráficas nos RIs brasileiros para visualização de indicadores. Tal conclusão corrobora com a percepção de que a representação mais comum de indicadores se dá em formato de índices, ou seja, é possível relacionar, por exemplo, o campo autor com o ano ou com descritores, entre outras relações. No entanto, ao seguir esse modo, a compreensão do “todo” é prejudicada.

Portanto, a interpretação mais abrangente de indicadores relacionais complexos exige representações gráficas dinâmicas, nas quais seja possível incluir e excluir propriedades bibliográficas, além de movimentá-las a partir de um gráfico em rede dinâmico de coocorrência.

Uso de indicadores relacionais complexos no ensino: o exemplo do Crai/Fousp

O Centro de Recursos para Aprendizagem e Investigação da Faculdade de Odontologia da Universidade de São Paulo (Crai/Fousp) configura um modelo inovador de métodos pedagógicos com ideais iniciados na Declaração de Bolonha, em 1999. Nesse sentido, o objetivo é tornar mais fácil para professores, pesquisadores e estudantes as atividades de aprendizagem, de formação, de gestão e de resolução de problemas, sejam técnicos, metodológicos ou de conhecimento, capacitando-os mais plenamente ao acesso e uso da informação (Novelli; Ramos; Pestana, 2016).

Com o conceito de aprendizagem baseado no “aprender a aprender” e apoiado nas tecnologias da informação, o Crai/Fousp é um ambiente integrado de laboratórios experimentais localizado na biblioteca com atividades voltadas aos alunos de graduação (Novelli; Ramos; Pestana, 2016). Uma de suas atividades é fornecer informação estratégica para alunos e professores, a partir da análise da produção científica da faculdade. Sendo assim, uma das análises bibliométricas mais importantes utilizadas é a de coocorrência com representação em rede, conforme Figura 3.

Figura 3 – Rede de descritores, orientadores e departamentos, a partir de teses defendidas nos últimos cinco anos na Fousp



Fonte: Novelli, Ramos e Pestana (2016).

A representação permite visualizar quais professores e departamentos trabalham com determinada temática e/ou quais temáticas importantes são menos pesquisadas. Tais informações estratégicas podem auxiliar o aluno no momento da escolha de tema de pesquisa e na escolha de seu futuro orientador.

Como visto, o projeto Crai/Fousp pode servir de exemplo de uso de gráficos em rede para diversas finalidades de ensino. Além disso, outros benefícios são observados, por exemplo: reduzindo a complexidade dos dados, ajudando a reconhecer as propriedades existentes e, conseqüentemente, auxiliando a detecção de erros aparentes; e englobando a representação em pouco conteúdo, na qual a consciência cognitiva é expandida, dentre outros (Segundo *et al.*, 2021).

Contudo, sabe-se que os RIs tradicionais são limitados nas representações gráficas de seus indicadores e que existe uma extensão/sistema que aprimora essa funcionalidade, conhecido como Current Research Information System (Cris) ou Research Information Management System (RIMS), ou seja, um sistema de gerenciamento de informação de pesquisa.

Current Research Information System

O Cris é um sistema de apoio à gestão institucional, que tem como núcleo central o RI, e congrega sistemas internos – institucionais – e externos para fornecer informações estratégicas de toda a atividade de pesquisa na

instituição – insumos, produtos e resultados; ou seja, transcende as funcionalidades dos RIs, considerando a complexidade e as relações do ambiente de pesquisa (Costa; Shintaku; Silveira, 2017; Sales; Sayão, 2015).

Segundo uma das principais iniciativas na área, o DSpace-Cris (2023, p. 1, tradução nossa) – mantida pela 4Science –, as finalidades mais conhecidas do Cris são:

- avaliar o desempenho da pesquisa;
- encontrar novas oportunidades de financiamento;
- evitar duplicação de atividades;
- analisar tendências;
- localizar novos contatos – especialistas;
- identificar novos mercados para produtos de pesquisa;
- disseminar pesquisas para apoiar o acesso aberto ao conhecimento científico.

Ao integrar sistemas internos e externos, o Cris não só organiza informações científicas dispersas, como também apresenta todo o ciclo de vida da atividade de pesquisa de uma instituição, relacionando financiamentos, projetos, produção científica, teses e dissertações, equipamentos, dados de pesquisa etc., a partir de um perfil de cada professor/pesquisador – ver exemplo da Figura 4.

Figura 4 – Exemplo de perfil de autor

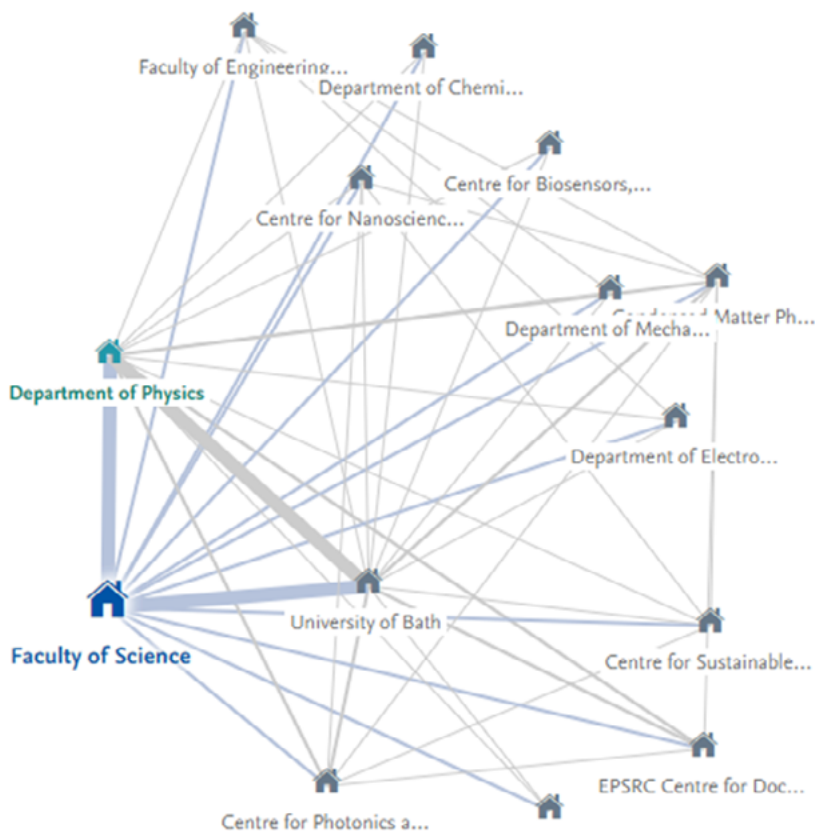


Fonte: University of Bath (c2024a).

Na Figura 4, é possível observar o perfil de uma pesquisadora da Universidade de Bath, que utiliza o Cris denominado Pure, da Elsevier, no qual apresenta na parte superior os conteúdos/sistemas da instituição: “Profiles”, “Research Units”, “Research outputs”, “Projects”, “Datasets” etc. Na parte de baixo, apresenta a relação do perfil da autora com outros conteúdos/sistemas, inclusive a sua “Network”.

Desse modo, é possível observar todo o histórico da atividade do pesquisador, além de obter análises de diversas relações existentes a partir da sua produção científica, da faculdade, do departamento, dos grupos de pesquisas etc., visualizadas em representações gráficas dinâmica de rede, conforme Figura 5 a seguir.

Figura 5 – Exemplo de gráfico dinâmico de rede de faculdades e departamentos da Universidade de Bath



Fonte: University of Bath (c2024b).

A partir do gráfico, é possível visualizar a relação – coautoria da produção científica – entre faculdades e departamentos e sua intensidade – espessura das linhas. É importante frisar que tal relação pode ser estabelecida a partir de outras propriedades bibliográficas, como, por exemplo: autores e departamentos; palavras-chave e autores etc.

Ao considerar uma breve história, embora o conceito e as iniciativas internacionais de criação de sistemas de gerenciamento da informação de pesquisa remontem da década de 1950 e, no Brasil, da década de 1980 (Barreto, 1981 *apud* Lyra, 2016), a informatização e evolução e a especialização das ciências, notadamente o crescimento exponencial da produção

científica, trouxeram novos desafios de organização da informação científica e, com isso, o surgimento de sistemas mais complexos com uso, por exemplo, de inteligência artificial, *web* semântica, ontologias etc.

Um sistema nesses moldes – e semelhante ao apresentado no exemplo acima da Universidade de Bath – foi aplicado no Brasil em nível institucional na Universidade Federal do Rio de Janeiro, conforme relato de experiência do projeto piloto publicado por Amaro (2017), que objetivou apresentar um diagnóstico de migração de dados do sistema DSpace – utilizado no RI – para o DSpace-Cris.

O DSpace-Cris, assim como o DSpace para RI, é também uma iniciativa internacional, customizado e disponibilizado pelo IBICT; sendo, segundo Costa, Shintaku e Silveira (2017, p. 251), “uma versão estendida do software livre, do código aberto, DSpace, de forma a atender as características de sistemas de informação CRIS” e mantendo as funcionalidades de um RI.

Em nível nacional, portanto mais amplo que os Cris institucionais, tem-se a implementação formal do projeto de pesquisa do Ecossistema de Informação da Pesquisa Científica Brasileira (BrCris) em 2020 (Segundo *et al.*, 2021) e mais recentemente da disponibilização da sua versão beta de iniciativa do IBICT. Segundo Pinto e demais autores (2021, p. 3), o BrCris é um ecossistema que permite a integração, por meio da interoperabilidade de dados, de diferentes sistemas nacionais, tais como: “agências de financiamento, bases de dados científicas e tecnológicas, repositórios institucionais, plataformas governamentais, infraestrutura de pesquisa, bases de teses e dissertações entre outros atores”.

É possível constatar a amplitude da adoção de sistemas de gerenciamento de pesquisa pela variedade de *softwares* disponíveis, tais como os mencionados anteriormente: DSpace-Cris da 4Science, Vivo, Pure da Elsevier, além do Converis da Clarivate e soluções próprias. Ademais, iniciativas internacionais de implementação, total de 339 em fevereiro de 2023, podem ser verificadas no *site* da The International Organisation for Research Information (euroCris), que é uma “associação internacional sem fins lucrativos fundada em 2002 com o objetivo de reunir especialistas em informação de pesquisa em geral e sistemas de informação de pesquisa (CRIS) em particular”.

Considerações finais

O sistema Cris foi apresentado neste trabalho, não somente como uma aprimoração/extensão/integração de RI no que se refere à integração entre sistemas internos/externos e gestão/avaliação de atividades de pesquisa de uma instituição, mas, principalmente, por sua função específica de permitir representações gráficas de redes de relacionamentos dinâmicas, necessárias para compreensão de indicadores complexos e seus possíveis usos no ensino, pesquisa, gestão e recuperação contextualizada da informação científica.

Sabe-se que os principais objetivos dos RIs são preservar e aumentar a visibilidade e uso da produção científica de uma instituição por meio da disponibilização de documentos em acesso aberto. Contudo, ainda é menor o número de documentos disponíveis se comparado à quantidade de registros bibliográficos nesses recursos, sendo uma das causas as restrições de direitos autorais de editoras comerciais.

Diante desse problema, considera-se importante utilizar, como valor agregado, tais registros bibliográficos para implementar indicadores relacionais complexos, que possibilitem analisar mais profunda e detalhadamente a produção científica. Apresentamos aqui, como exemplo, o uso dessas análises para fins de ensino no caso do Crai/Fousp.

Para tanto, a integração/interoperabilidade de sistemas internos e externos é fundamental, considerando ainda os RIs e perfis de pesquisadores como fontes principais. Diante disso, o Cris reúne todas essas características ampliando o relacionamento destes na produção científica.

Referências

AMARO, J. de M. *Diagnóstico de migração de dados do Repositório Institucional da UFRJ (Pantheon) para um repositório baseado em DSpace-CRIS*. 2017. Trabalho de Conclusão de Curso (Bacharelado em Biblioteconomia e Documentação) – Instituto de Arte e Comunicação Social, Universidade Federal Fluminense, 2017. Disponível em: <https://app.uff.br/riuff/bitstream/handle/1/24478/JAIRO%20DE%20M.%20AMARO%20%282017%29.pdf?sequence=1&isAllowed=y>. Acesso em: 13 jan. 2023.

AZADBAKHT, E.; SCHULTZ, T. At the Click of a Button: Assessing the User Experience of Open Access Finding Tools. *Information Technology and Libraries*, [s. l.], v. 39, n. 2, p. 1-13, June 2020.

CONFEDERATION OF OPEN ACCESS REPOSITORIES. *Next Generation Repositories: Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group*. [Lincoln]: COAR, 2017. Disponível em: <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1065&context=scholcom>. Acesso em: 22 set. 2020.

COSTA, L. R. da; SHINTAKU, M.; SILVEIRA, L. A. da. DSpace CRIS. In: VECHIATO, F. L. et al. (org.). *Repositórios digitais: teoria e prática*. Curitiba: EDUTFPR, 2017. p. 245-264.

DHAKAL, K. Unpaywall. *Journal of the Medical Library Association*, [s. l.], v. 107, n. 2, p. 286-288, Apr. 2019. DOI 10.5195/jmla.2019.650. Disponível em: <https://jmla.pitt.edu/ojs/jmla/article/view/650>. Acesso em: 4 ago. 2020.

DIMENSIONS. *Dimensions*. London, 2023. Disponível em: <https://www.dimensions.ai/>. Acesso em: 3 fev. 2023.

DSPACE-CRIS. *DSpace-Cris Home*. Kansas, 2023. Disponível em: <https://wiki.lyrasis.org/display/DSpaceCRIS/DSpace-CRIS+Home>. Acesso em: 8 fev. 2023.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. *Portaria CTA nº 1/2019*. Institui a “Política para Acesso Aberto às publicações Resultantes de Auxílios e Bolsas Fapesp. São Paulo: Fapesp, 2019. Disponível em: <https://fapesp.br/12632/portaria-cta-no-012019>. Acesso em: 8 fev. 2023.

GARCÍA-PEÑALVO, F. J. El futuro de los repositorios institucionales. *Education in the Knowledge Society*, [s. l.], v. 18, n. 4, p. 7, 2017. Disponível em: <https://revistas.usal.es/tres/index.php/eks/article/view/eks2017184719/17982>. Acesso em: 8 fev. 2023.

LUTHER, J. The evolving institutional repository landscape. *Choice: Current Reviews for Academic Libraries*, [s. l.], v. 55, n. 7, p. 796, 2018.

LYRA, M. C. T. C. *Sistema de informação sobre pesquisa baseado no modelo CRIS: conceitos, tecnologias, padrões e aplicações*. 2016. Dissertação (Mestrado em Biblioteconomia) – Universidade Federal do Estado

do Rio de Janeiro, Rio de Janeiro, 2016. Disponível em: http://www.repositorio-bc.unirio.br:8080/xmlui/bitstream/handle/unirio/11934/MAURA_CRISTINA_TUPINQUIM_CAVALCANTI_LYRA.pdf?sequence=1&isAllowed=y. Acesso em: 8 fev. 2023.

MELEIRO, R. Evolució i nova generació dels repositoris. *Blok de Bid*, [s. l.], 28 fev. 2018. Disponível em: <http://www.ub.edu/blokdebid/ca/node/850>. Acesso em: 16 set. 2020.

MOREIRA, J. M. *et al.* Integrating a national network of institutional repositories into the national/international research management ecosystem. *Procedia Computer Science*, [s. l.], v. 106, p. 146-152, 2017.

NOVELLI, M. D.; RAMOS, L. M. S. V. C.; PESTANA, M. C. *Concepts of social networking in dentistry: pedagogical models and teaching*. London: LAP LAMBERT Academic Publishing, 2016.

OPENDOAR. *Directory of Open Access Repositories*. Nottingham: University of Nottingham: Lund University, 2023. Disponível em: <https://v2.sherpa.ac.uk/opendoar/search.html>. Acesso em: 8 fev. 2023.

PINTO, A. L. *et al.* BrCris como um sistema de recomendação científico-tecnológica. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 21., 2021, Rio de Janeiro. *Anais [...]*. Rio de Janeiro: IBICT/UFRJ, 2021. p. 1-15. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/viewFile/335/456>. Acesso em: 8 fev. 2023.

REIS, J. E. dos; SPINOLA, A. T. P.; AMARAL, R. M. do. Incipiência da visualização de indicadores bibliométricos e altmétricos nos Repositórios Institucionais brasileiros. *Em Questão*, Porto Alegre, v. 23, p. 213-234, 2017. Edição especial 5º EBBC. DOI 10.19132/1808-5245230.213-234. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/68070/40190>. Acesso em: 7 ago. 2020.

SALES, L. F.; SAYÃO, L. F. Ciberinfraestrutura de informação para a pesquisa: uma proposta de arquitetura para integração de repositórios e sistemas Cris. *Informação & Sociedade: Estudos*, João Pessoa, v. 25, n. 3, p. 163-184, set./dez. 2015. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/23998/14535>. Acesso em: 13 jan. 2023.

SEGUNDO, W. *et al.* Uma plataforma para o tratamento e integração de dados científicos. In: CONGRESSO LATINO-AMERICANO DE

SOFTWARE LIVRE E TECNOLOGIAS ABERTAS, 18., 2021, Foz de Iguaçu. *Anais [...]*. Foz de Iguaçu: Sociedade Brasileira de Computação, 2021. p. 108-113. Disponível em: <https://sol.sbc.org.br/index.php/latinoware/article/view/19913/19741>. Acesso em: 8 fev. 2023.

SHEARER, K. Spotlight on Open Access repositories. *Online Researcher*, v. 41, n. 1, p. 63-65, 2017.

TAY, C. H. A. Rethinking institutional repositories. *Online Researcher*, [s. l.], v. 41, n. 12, p. 10-15, 2017. Disponível em: https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=1104&context=library_research. Acesso em: 7 ago. 2020.

UNIVERSITY OF BATH. *Andrea Abbas, Professor of the Sociology of Higher Education*. Bath, c2024a. Disponível em: <https://researchportal.bath.ac.uk/en/persons/andrea-abbas>. Acesso em: 11 jan. 2023.

UNIVERSITY OF BATH. *Department of Physics*. Bath, c2024b. Disponível em: <https://researchportal.bath.ac.uk/en/organisations/department-of-physics/network/>. Acesso em: 11 jan. 2023.

WU, M. The future of institutional repositories at small academic institutions: analysis and insights. *D-Lib Magazine*, [s. l.], v. 21, n. 9/10, Sept./Oct. 2015. DOI 10.1045/september2015-wu. Disponível em: <https://www.dlib.org/dlib/september15/wu/09wu.html>. Acesso em: 7 ago. 2020.

YOUNGBLOOD, M.; LAHTI, D. A bibliometric analysis of the interdisciplinary field of cultural evolution. *Palgrave Communications*, v. 4, n. 120, p. 1-9, 2018. DOI 10.1057/s41599-018-0175-8. Disponível em: <https://www.nature.com/articles/s41599-018-0175-8>. Acesso em: 8 fev. 2023.

Parte III

Tecnologias

CAPÍTULO 8

O registro de fontes de dados cientométricos¹

Grischa Fraumann

Svantje Lilienthal

Christian Hauschke

Introdução

O número e a variedade de fontes de dados cientométricos *on-line* estão aumentando constantemente, o que é um forte contraste com as poucas fontes que existiam há alguns anos. Essas fontes mais tradicionais são amplamente discutidas na literatura, como os bancos de dados bibliográficos Scopus (Baas *et al.*, 2020), Web of Science (WoS) (Birkle *et al.*, 2020) e Google Scholar (Gusenbauer, 2019). Na era digital, há necessidade de uma visão geral abrangente sobre fontes convencionais e emergentes – dados abertos – (Waltman; Larivière, 2020), o que foi expresso em várias iniciativas nacionais e internacionais, como o Manifesto de Leiden para Métricas de Pesquisa (Hicks *et al.*, 2015).

Descrevemos um registro público colaborativo aberto de tais fontes, o Registry of Scientometric Data Sources – ou Registro de Fontes de Dados Cientométricos –, que fornece essa visão geral e para o qual os usuários da comunidade acadêmica e de outras áreas são convidados a contribuir (Lilienthal, 2019). Para isso, formulamos a seguinte pergunta de pesquisa: “quais fontes de dados cientométricos estão disponíveis para fins de pesquisa?”. Primeiro, avaliamos as fontes de dados internamente como parte do projeto de pesquisa “Reference Implementation for Open Scientometric Indicators (Rosi)” (Hauschke; Cartellieri; Heller, 2018) e depois decidimos

¹ Esta pesquisa foi financiada pelo German Federal Ministry of Education and Research (BMBF) sob o número de concessão 01PU17019. O link para o *preprint* pode ser acessado em: <https://arxiv.org/pdf/2311.06157>.

compartilhar nossas descobertas na forma do Registro de Fontes de Dados Cientométricos (RFDS)².

O registro é um serviço para pesquisadores especializados em pesquisas cientométricas, pesquisas institucionais, bibliotecários, profissionais, formuladores de políticas, estudantes e quaisquer interessados em cientometria que gostariam de decidir quais fontes de dados usar em projetos de pesquisa, pesquisar ou adicionar novas fontes de dados. Adotamos uma abordagem de *crowdsourcing* que também foi usada em iniciativas semelhantes (Heibi; Peroni; Shotton, 2019; Open Science Monitor, 2019). O registro é, portanto, uma nova infraestrutura de pesquisa aberta. Juntamente com o acesso aberto a publicações e dados de pesquisa, as infraestruturas de pesquisa estão sendo abertas cada vez mais para a sociedade em geral. Por exemplo, as infraestruturas de pesquisa do Joint Research Centre (JRC) da Comissão Europeia foram abertas por meio de chamadas para participação da comunidade acadêmica e do público em geral (European Commission, 2020).

Histórico e revisão da literatura

O surgimento da cientometria está intimamente ligado à disponibilidade de fontes de dados legíveis por máquina. Nos primórdios da disciplina, era necessário agregar manualmente os dados das próprias publicações ou das bibliografias. Um exemplo importante desse último caso é o trabalho feito por Lotka (1926) para “The frequency distribution of scientific productivity” – A distribuição de frequência da produtividade científica –, em que ele usou o índice de pessoas no Chemical Abstracts 1907-1916, e no *Geschichtstafeln der Physik* de Auerbach, para identificar pesquisadores proeminentes de acordo com o número de publicações. Embora seu método tenha sido bem-sucedido, ele era naturalmente muito tedioso.

A primeira ferramenta a fornecer acesso eletrônico conveniente às informações de citação foi o Science Citation Index (SCI), introduzido em 1955 na revista *Science* e lançado em 1964 (Garfield, 2007). Além disso, o Journal Citation Reports (JCR) foi lançado em 1975 e as classificações do

2 Disponível em: <https://labs.tib.eu/rosi>.

fator de impacto foram estabelecidas. Em termos simples, o fator de impacto calcula o número médio de citações por artigo de um determinado periódico, geralmente com base nos últimos dois anos. O fator de impacto recebeu uma quantidade considerável de críticas que não são o foco deste artigo (Lariviere; Sugimoto, 2019). O SCI exibe o número de publicações – ou seja, o Source Author Index – e as citações de pesquisadores individuais – ou seja, o Citation Index. No JCR, as citações são ordenadas de acordo com o nome do periódico. Essas duas medidas também podem ser estendidas para o nível institucional e/ou nacional. O banco de dados bibliográfico WoS foi descrito por Garfield (2007) como a versão eletrônica do SCI, entre outros.

A próxima etapa evolutiva foi a disponibilidade de dados cientométricos por meio de Interfaces de Programação de Aplicativos – Application Programming Interfaces – (APIs). Em termos simples, uma API “[...] é uma interface que fornece acesso programático à funcionalidade e aos dados do serviço em um aplicativo ou banco de dados” (Gartner, 2020). As APIs estão se tornando cada vez mais comuns em ambientes de pesquisa distribuídos. Por exemplo, uma API permite que os pesquisadores recuperem sempre os dados mais recentes de uma interface e possibilita a integração em outro sistema (National Information Standards Organization, 2016). A disponibilidade de APIs também é uma abordagem que pode ser estudada ao investigar fontes de dados. Outro aspecto é a disponibilidade de documentações que descrevem o uso de APIs (National Information Standards Organization, 2016).

As APIs são amplamente usadas para fins de pesquisa, e até mesmo sistemas que normalmente exigem uma assinatura – como o Scopus – fornecem esse serviço a pesquisadores ou outros usuários (National Information Standards Organization, 2016). Obviamente, serviços públicos, como a Wikipédia, também fornecem API para recuperar dados (National Information Standards Organization, 2016). No entanto, alguns bancos de dados bibliográficos conhecidos, como o Google Scholar, não fornecem acesso por meio destas (Balstad; Berg, 2020). Há o risco de que algumas dessas interfaces sejam descontinuadas, por exemplo, a API do *site* de rede profissional LinkedIn tem sido usada desde 2013 para estudos altmétricos (Altmetric.com, 2021a), por exemplo, para investigar como os resultados de

pesquisa foram compartilhados em perfis públicos e em grupos. No entanto, a API foi descontinuada em 2014. Os provedores de dados podem fornecer uma interface de envio para permitir que as fontes de dados enviem informações ao vivo para o sistema ou para várias instâncias (National Information Standards Organization, 2016). A atualidade das respostas da API pode depender do provedor de dados (National Information Standards Organization, 2016).

Alguns desses provedores oferecem o chamado acesso *premium* por meio de um contrato de nível de serviço que garante determinadas funções, como tempos de resposta ou número de chamadas em um intervalo de tempo especificado (National Information Standards Organization, 2016). Essa limitação foi introduzida, por exemplo, pelo Crossref Event Data (2021), um serviço do consórcio de editores Crossref que rastreia vários eventos *on-line*, como as menções de resultados de pesquisa nas mídias sociais e na Wikipédia. O formato retornado de uma chamada de API é geralmente o JavaScript Object Notation (JSON) (National Information Standards Organization, 2016). Em resumo, as APIs permitem a aquisição de dados cientométricos de vários sistemas por meio de interfaces padronizadas, o que promove a pesquisa e o desenvolvimento em ambientes distribuídos e contribui para a bibliometria aberta, conforme descrito por Stuart (2018). As APIs também podem ser um recurso importante para a realização de estudos altmétricos. O campo da webometria ganhou força no início dos anos 2000, com a análise de várias novas fontes de dados *on-line* por cientometristas. A webometria é definida como “o estudo quantitativo de fenômenos relacionados à Web” (Thelwall; Vaughan; Björneborn, 2005). Outro marco nos estudos de mídia social da ciência (Costas, 2017) e da *web* em geral foi o Manifesto Altmetrics de 2010 (Priem *et al.*, 2010), enquanto a visibilidade na *web* de resultados acadêmicos já havia sido explorada anteriormente (Aaltojärvi *et al.*, 2008). O objetivo do Manifesto Altmetrics era introduzir a altmetria como um complemento aos indicadores bibliométricos convencionais devido à quantidade crescente de informações científicas disponíveis *on-line* e medir como os usuários se envolvem com esses dados. Esse desenvolvimento estava relacionado ao aumento da mídia social e das plataformas colaborativas *on-line*. Independentemente desses avanços, não é

possível obter acesso totalmente aberto às informações cientométricas. Para isso, dentre outras, as seguintes perguntas precisam ser respondidas:

- onde os usuários podem obter os dados?
- quais restrições existem para a reutilização gratuita dos dados para qualquer finalidade?
- como os dados podem ser recuperados automaticamente?

A abertura dos dados cientométricos é um tópico que recebe cada vez mais atenção da comunidade acadêmica (Waltman; Larivière, 2020), por exemplo, em relação à abertura de listas de referências (Mugnaini *et al.*, 2021; Sugimoto; Murray; Larivière, 2018) ou à disponibilidade de resumos de publicações (Tay; Kramer; Waltman, 2020), porque a pesquisa cientométrica só é possível se o acesso a esses dados for garantido. Conforme mencionado anteriormente, as fontes de dados abertas são necessárias para facilitar a pesquisa cientométrica, já que os pesquisadores precisam de fontes de abertas para garantir acessibilidade, transparência dos resultados e replicabilidade da pesquisa. Isso é postulado por muitas iniciativas, por exemplo, pelos fundadores da revista *Quantitative Science Studies* (Waltman; Larivière, 2020). A avaliação da pesquisa por instituições executoras de pesquisa, organizações de financiamento de pesquisa, agências de avaliação de pesquisa, pesquisadores e outras partes interessadas deve se basear em dados abertos e indicadores transparentes, de acordo com várias iniciativas internacionais, como a Declaração de São Francisco sobre Avaliação de Pesquisa (Dora) (Cagan, 2013), o Manifesto de Leiden para Métricas de Pesquisa (Hicks *et al.*, 2015) e os Princípios de Hong Kong para Avaliação de Pesquisadores (Moher *et al.*, 2020), bem como iniciativas nacionais que avaliam essa abertura – por exemplo, o Ministério da Educação e Cultura da Finlândia, 2016. Há vários casos de uso de fontes de dados abertos para pesquisa cientométrica, dos quais apresentamos uma seleção:

- comparação das contagens de citações da mesma publicação fornecidas por várias fontes de dados bibliométricos;

- correspondência de publicações por meio de Digital Object Identifiers, ou Identificadores de Objetos Digitais (DOIs), entre várias fontes de dados, por exemplo, para enriquecer os metadados bibliográficos;
- estimativa da cobertura de disciplinas acadêmicas em diferentes fontes de dados bibliométricos;
- investigação do número de publicações de acesso aberto em diferentes fontes de dados bibliométricos;
- realização de estudos cientométricos, por exemplo, sobre a recepção de resultados de pesquisa em várias fontes de dados.

Nosso estudo faz parte de um projeto maior sobre desenvolvimento iterativo de *software* com base em um *design* amigável ao usuário (Hauschke; Cartellieri; Heller, 2018), o projeto Rosi.

Iniciativas relacionadas

Além do projeto Rosi, o Metrics Project lançou o Social Media Registry³ – Registro de Mídia Social –, que foi levado em consideração neste estudo. Enquanto o último se concentrou apenas em fontes de mídia social, o Registro de Fontes de Dados Cientométricos (RFDC) inclui outras fontes de dados cientométricos, como as que contêm dados bibliométricos ou de patentes. Fizemos uma revisão da literatura de registros semelhantes e da literatura relacionada. Os registros de fontes e bancos de dados são uma prática comum no meio acadêmico, por exemplo, para descrever bancos de dados bibliográficos nacionais (Sile *et al.*, 2017) e serviços para Identificadores Persistentes (PIDs) de forma estruturada (PID Services Registry, 2020). Houve iniciativas semelhantes, como o Metrics Project, o projeto European Union Relevant, Inclusive, Trusted and Open Indicators for Research and Innovation Policy (Eurito), o projeto Data4Impact, o Open Science Monitor (OSM) e o Metrics Toolkit⁴.

O Metrics Project concentrou-se especialmente em métricas de mídia social para as ciências sociais e incluiu estudos de usuários de pesquisadores

3 Disponível em: <https://www.leibniz-openscience.de/lra-projects/metrics/>

4 Disponível em: https://www.metrics-toolkit.org/metrics/journal_impact_factor/.

(Lemke *et al.*, 2019). Um dos resultados do projeto foi o desenvolvimento do registro de mídias sociais mencionado acima. O consórcio Eurito realizou uma auditoria de novas fontes de dados que poderiam ser usadas como evidência para políticas de pesquisa e inovação (Tippet; Blind; Parraguez Ruiz, 2018) e divulgou uma visão geral dos casos-piloto e das possíveis fontes de dados (Eurito Consortium, 2018). O consórcio Data4Impact investigou fontes de dados para desenvolver “um sistema de apoio ao cálculo automático de indicadores sobre o impacto das políticas de ciência, tecnologia e inovação” (Manghi; Atzori, 2018). Como parte do desenvolvimento do OSM, as fontes de dados foram avaliadas e a comunidade acadêmica foi solicitada a fornecer *feedback* (Open Science Monitor, 2019).

O Metrics Toolkit oferece uma visão geral dos indicadores cientométricos e das fontes de dados relacionadas. Ele é apoiado pelo provedor de dados Altmetric, pela Universidade de Indiana e pela iniciativa de comunicações acadêmicas FORCE11. O Metrics Toolkit pode ser visto como uma iniciativa semelhante que foi desenvolvida antes do registro e fornece seus dados sob uma licença internacional Creative Commons Attribution (CC BY) 4.0. Em contraste com o Metrics Toolkit, não formamos um conselho consultivo porque gostaríamos que a comunidade contribuísse com as fontes de dados. Além disso, não fornecemos informações detalhadas sobre indicadores, por exemplo, indicadores baseados em periódicos, como o Journal Impact Factor⁵. Nosso foco está em discriminar as próprias fontes de dados.

Além do Metrics Toolkit, o Snowball Metrics é uma iniciativa que fornece informações sobre métricas desenvolvidas pela comunidade acadêmica⁶. Além disso, vários grupos de especialistas de alto nível da União Européia (UE) discutiram fontes de dados como parte de seus relatórios finais (Cabello Valdes *et al.*, 2017; European Commission, 2018; Guédon *et al.*, 2019; Von Schomberg, 2019; Wilsdon *et al.*, 2017).

Há também um projeto que realizou uma pesquisa sobre novas fontes de dados para comunicação acadêmica (Bosman; Kramer, 2016) e uma iniciativa com foco em fontes de dados altmétricos com base na contribuição

5 Disponível em: <https://www.metrics-toolkit.org/journal-impact-factor/>.

6 Disponível em: <https://snowballmetrics.com/>.

de grupos de trabalho de especialistas, que foi conduzida pela Organização Nacional de Padrões de Informação dos Estados Unidos da América (National Information Standards Organization, 2016). Por fim, levamos em conta as descrições das fontes de dados pelos provedores de dados altmétricos (Altmetric.com, 2021b; Plum Analytics, 2019).

O projeto Rosi

O objetivo do projeto Rosi é desenvolver uma visualização de indicadores cientométricos com base em fontes de dados abertas, levando em conta as necessidades do usuário. Não desenvolvemos novos indicadores, mas reutilizamos aqueles de fontes de dados disponíveis abertamente. Os estudos de usuários foram realizados em entrevistas qualitativas e *workshops* com grupos focais para reunir as necessidades e preocupações dos pesquisadores como possíveis usuários da visualização.

Por fim, foram realizadas entrevistas para testar a usabilidade do *software*. Como parte do projeto, avaliamos as fontes de dados que são adequadas para gerar indicadores cientométricos. Essa avaliação levou ao desenvolvimento do registro. O projeto está inserido na linha de financiamento Quantitative Research on the Science Sector – Pesquisa Quantitativa sobre o Setor Científico – do Ministério Federal Alemão de Educação e Pesquisa (BMBF) e o objetivo é colaborar com projetos relacionados.

Como favorecemos as fontes de dados que fornecem ou pretendem desenvolver uma API, a maioria das fontes de dados no RFDC atende a esse critério. Avaliamos as fontes de dados cientométricos para apoiar a comunidade especialista no desenvolvimento de sua própria infraestrutura e para fornecer informações adicionais sobre essas fontes de dados cientométricos abertas.

Registro de Fontes de Dados Cientométricos (RFDC)

Com o fim de avaliar as fontes de dados definimos um esquema de metadados e uma estrutura de dados – ou seja, fontes, indicadores, categorias, entidades – para uma descrição técnica. Seleccionamos as fontes de dados

com base em critérios, como relevância para estudos cientométricos internacionais e APIs planejadas ou já existentes, SPARQL Endpoints ou outras interfaces para recuperar ou fazer *download* de dados.

Para publicar o registro, desenvolvemos um *site* que exhibe as fontes de dados, o esquema de metadados, as metas do registro e as informações adicionais sobre o projeto. Uma versão inicial do registro foi apresentada em vários *workshops* e conferências, e anunciado em um artigo de *blog* (Lilienthal, 2019) e nas mídias sociais. Discutimos o registro com os parceiros do projeto, incluindo o projeto Metrics e o projeto Open Access Publication at Universities in Germany: Development and Influencing Factors (OAUNI)⁷, membros da comunidade cientométrica e administradores de pesquisa e gestores de pesquisa, obtendo assim um valioso *feedback* dos usuários sobre fontes de dados adicionais – incluindo fontes de dados não abertas – e possíveis melhorias no registro.

Nossa análise resultou em uma lista colaborativa e publicamente disponível de fontes de dados que podem ser usadas para recuperar informações cientométricas: o RFDC. Nesta seção, descreveremos o esquema de metadados e os recursos do registro, que apresentam uma visão geral das fontes de dados e visualizações detalhadas de cada uma delas – com opção de edição –, e fornece uma visão geral técnica das fontes de dados e suas conexões.

Esquema de metadados

As entradas no registro são estruturadas de acordo com o seguinte esquema de metadados (Quadro 1).

Quadro 1 - Esquema de metadados no RFDC

No.	Campo de metadados	Definição
1	Nome	O nome da fonte de dados.

7 Disponível em: <https://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/oauni/>.

2	Link	O Localizador Uniforme de Recursos (URL) para a página principal da fonte de dados.
3	Link para o logotipo	Um <i>link</i> (persistente) para uma imagem.
4	Descrição	Uma descrição da própria fonte de dados.
5	Categoria	Uma categoria que descreve a fonte de dados – plataforma colaborativa, rede social <i>on-line</i> , repositório, agregador de dados, outro.
6	Entidade	Uma entidade que pode ser descrita por essa fonte de dados – trabalho, pessoa, organização, evento, projeto.
7	Licença	A licença dos dados.
8	Interface	O <i>link</i> para uma API ou outra interface.
9	Tipo de interface	O tipo de interface – API REST, API HTTP, API HTTPS, SPARQL Endpoint, outro.
10	Documentação	O <i>link</i> para a documentação da interface.
11	Formato de dados	O formato de dados da resposta da API – nenhum, JSON, XML, outro.
12	Comentário	Uma área de texto para qualquer forma de comentário.

Fonte: Leibniz Information Centre for Science and Technology ([20--a]).

Nota: Uma fonte de dados também pode descrever várias entidades, mas somente uma entidade pode ser selecionada para cada fonte de dados no registro.

Visão geral e detalhada das fontes de dados

O registro fornece uma visão geral de todas as fontes de dados na página inicial do aplicativo com algumas informações básicas, incluindo o nome, uma descrição e um *link* para a fonte de dados – consulte a Figura 3 e a página do GitHub⁸ para ver o *status* atual das entradas de dados.

8 Disponível em: <https://github.com/TIBHannover/rosi-registry>.

Figura 1 – Página inicial do RFDC⁹

Registry of Scientometric Data Sources

Home Technical Overview Dataflow About

This registry describes data sources for scientometric information. It is maintained by the ROSI project. [Read more.](#)

This is a work in progress. You are invited to help filling the registry! Contact us to get write access: [rosi.project\(at\)zb.eu](mailto:rosi.project(at)zb.eu).

Add new data source

Show entries Search:

Name	Description	
Altmetric Details Page API	"The open Altmetric Details Page API allows rate-limited querying of Altmetric metrics for research outputs."	
BASE	"BASE is one of the world's most voluminous search engines especially for academic web resources. BASE provides more than 120 million documents from more than 6.000 sources. You can access the full texts of about 60% of the indexed documents for free"	
BibSonomy	Bibsonomy is a social bookmarking and publication sharing platform.	
Cobaltmetrics	"Cobaltmetrics crawls the web to index hyperlinks and persistent identifiers as first-class citations. We analyze a wide range of websites to reveal insightful links between documents." https://cobaltmetrics.com/	
COCI	"As of the most recent update on 4th July 2020, COCI contains metadata about 721.655,392 citations involving 58,876,621 bibliographic resources."	
CORE	"The world's largest collection of open access research papers."	
Crossref	"Crossref interlinks millions of items from a variety of content types, including journals, books, conference proceedings, working papers, technical reports, and data sets." https://en.wikipedia.org/wiki/Crossref	
Crossref Event Data	The Event Data service captures data on discussions about scholarly content in non-traditional places (online platforms for discussion, publication and social media) and acts as a hub for the storage and distribution of this data. The service provides a record of instances where research has been bookmarked, linked, liked, shared, referenced, commented on etc. beyond publisher platforms. For example, when datasets are mentioned on social media or referenced in Wikipedia.	
DataCite	DataCite assigns persistent identifiers (digital object identifiers, DOIs) to research data. This provides the opportunity to locate and cite research data, among others.	
dblp Computer Science Bibliography	"The dblp computer science bibliography is the on-line reference for bibliographic information on major computer science publications. It has evolved from an early small experimental web server to a popular open-data service for the computer science community. Our mission at dblp is to support computer science researchers in their daily efforts by providing free access to high-quality bibliographic meta-data and links to the electronic editions of publications." https://dblp.uni-trier.de/faq/WhatIs+dblp.html	

Showing 1 to 10 of 39 entries Previous 1 2 3 4 Next

Fonte: Leibniz Information Centre for Science and Technology ([20--b]).

Os metadados detalhados de cada fonte de dados podem ser recuperados em uma nova página do aplicativo. Nessa visualização, o registro oferece a possibilidade de pesquisar a literatura sobre a fonte de dados nos bancos de dados bibliográficos Connecting Repositories (Core)¹⁰ e Bielefeld Academic Search Engine (Base)¹¹. Um exemplo de uma visualização detalhada de uma fonte de dados, o Índice OpenCitations de citações DOI-to-DOI abertas do Crossref (Coci), é mostrado na Figura 2.

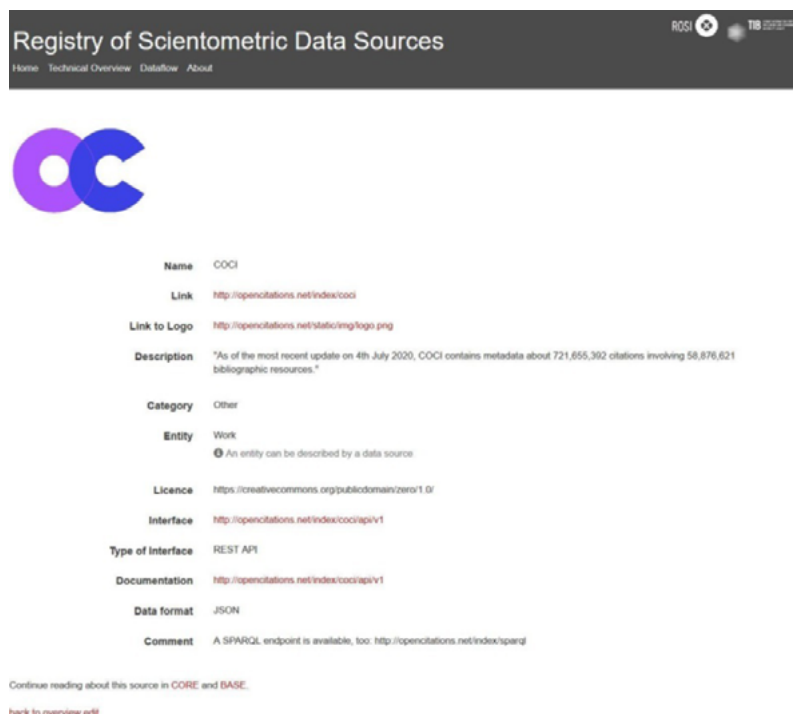
⁹ Esta página exibe informações básicas sobre as fontes de dados, como o nome, uma descrição e um *link* para a fonte de dados.

¹⁰ Disponível em: <https://core.ac.uk/>.

¹¹ Disponível em: <https://www.base-search.net/>.

Os usuários podem inserir os metadados de uma fonte de dados no registro seguindo um *link* de edição na visualização detalhada. Para rastrear a proveniência e garantir a integridade dos dados, eles precisam obter uma senha.

Figura 2 – Captura de tela da fonte de dados, Coci, e seus metadados no RFDC



Fonte: Leibniz Information Centre for Science and Technology ([20--c]).

Visão geral técnica de fontes de dados

Além da visão geral, uma página separada fornece uma visão geral técnica (Figura 3) que descreve os metadados de cada fonte de dados. Esses metadados são úteis para terceiros que gostariam de usar as fontes de dados para fins de pesquisa.

Figura 3 – Visão geral técnica das fontes de dados no RFDC

ROSI

TIB

Registry of Scientometric Data Sources

[Home](#) [Technical Overview](#) [Dataflow](#) [About](#)

This registry describes data sources for scientometric information. It is edited by the ROSI project. [Read more](#)

This is a work in progress. You are invited to help filling the registry! Contact us to get write access: rosi.project@tib.eu

Add new data source

Show 10 entries

Search:

ID	Name	Interface	Type of Interface	Documentation	Data Format
0	Directory of Open Access Journals (DOAJ)	https://doaj.org/api/v1/	API	https://doaj.org/api/v1/docs	JSON
1	dissemin	https://dev.dissemin.in/api.html	API	https://dissemin.readthedocs.io/en/latest/api.html	JSON
2	Crossref Event Data	https://api.eventdata.crossref.org	REST API	https://www.eventdata.crossref.org/guide/service/query-api/	JSON
3	OpenAIRE	http://api.openaire.eu	REST API	http://api.openaire.eu/	XML
4	NOA Scientific Image Search	https://noa.wp.hs-hannover.de/api.php?query=	API	http://noa.wp.hs-hannover.de/api-documentation.php	JSON
5	BASE	http://api.base-search.net/	API	http://www.base-search.net/about/download/base_interface.pdf	JSON
6	DataCite	https://api.datacite.org	REST API	https://support.datacite.org/docs/api	JSON
7	Wikipedia	https://en.wikipedia.org/w/api.php	API	https://de.wikipedia.org/wiki/Wikipedia:Technik/Datenbank/API	JSON
8	Publons	https://publons.com/api/v2/	REST API	https://publons.com/api/v2/	JSON
9	Open Patent Services (OPS)		REST API	http://documents.epo.org/projects/babylon/iponet.nsf/0/F3ECD0C915C9BCD6C125806003AA712?file/ops_v3_2_documentation_-_version_1.3.81_en.pdf	XML

Showing 1 to 10 of 39 entries

Previous 1 2 3 4 Next

Fonte: Leibniz Information Centre for Science and Technology ([20--d]).

Considerações finais

Trabalhar com o RFDC deixou claro que essa coleção de fontes de dados pode ser útil para aqueles que têm interesse em cientometria. No futuro, serão incluídos mais casos de uso para fontes de dados abertas, e pretendemos avaliar mais fontes de dados.

Além disso, gostaríamos de ampliar o esquema de metadados, o que poderia levar a um esquema de dados como base para um Protocolo de Transferência de Dados Cientométricos (SDTP), sugerido uma vez por Waltman (2019).

De modo geral, as fontes de dados abertos estão em constante mudança, e é por isso que algumas fontes podem ser adicionadas ao registro ou substituídas no futuro. Simplificando, “o trabalho [...] nunca está completo” (Wass, 2018) e as fontes precisam ser constantemente monitoradas

e discutidas (Zahedi; Costas, 2018). Até o momento, fornecemos apenas informações sobre uma seleção de fontes de dados e fazemos um apelo à comunidade cientométrica e a outras partes interessadas para que contribuam com o registro. As contribuições podem variar, por exemplo, desde a sugestão de fontes de dados adicionais e comentários sobre as fontes de dados existentes até a extensão do esquema de metadados.

Além disso, uma comparação da cobertura das fontes de dados (Visser; Van Eck; Waltman, 2021) está atualmente fora do escopo, mas seria um recurso benéfico do registro. Voltando à nossa pergunta de pesquisa inicial, ou seja, “quais fontes de dados cientométricos estão disponíveis para fins de pesquisa?”, descrevemos um processo para criar esse registro que permite *feedback* e contribuições de novas informações da comunidade acadêmica. Já recebemos *feedback* sobre o registro de alguns usuários, mas precisamos encontrar usuários que estejam dispostos a contribuir com informações sobre novas fontes de dados. Como as fontes de dados *on-line* são dinâmicas, com atualizações constantes e alterações ocasionais nos URLs, a atualidade e a sustentabilidade desse registro seriam mais facilmente garantidas com um envolvimento mais amplo da comunidade cientométrica. As baixas taxas de participação de outros usuários também podem ser observadas em outras solicitações de participação, por exemplo, para fornecer dados de citação para preencher índices de dados de citação abertos (Heibi; Peroni; Shotton, 2019).

No entanto, esperamos que o registro possa servir à comunidade acadêmica na realização de pesquisas cientométricas e se tornar um recurso para usuários que gostariam de saber mais sobre fontes de dados cientométricos. Em última análise, essa abordagem deve contribuir para a bibliometria e a cientometria abertas, fornecendo informações abrangentes sobre fontes de dados e como elas estão interconectadas.

Referências

AALTOJÄRVI, I. *et al.* Scientific Productivity, Web Visibility and Citation Patterns in Sixteen Nordic Sociology Departments. *Acta Sociologica*, [s. l.], v. 51, n. 1, p. 5-22, 2008. DOI 10.1177/0001699307086815. Disponível em:

https://journals.sagepub.com/doi/pdf/10.1177/0001699307086815?casa_token=uq7ayh4bKfQAAAAA;jNnVhnZdW-DsZ5wIZZFq8wG1SPr4p00ZADeQrJ11Bf0uk4-0gsfgMh9YpPC-1y6Rneaq87MMA8wujw. Acesso em: 5 out. 2024.

ALTMETRIC.COM. *Attention Sources coverage dates*. [S. l.]. 2021a. Disponível em: <https://help.altmetric.com/support/solutions/articles/6000240455-attention-sources-coverage-dates>. Acesso em: 5 out. 2024.

ALTMETRIC.COM. *Sources of Attention*. [S. l.], 2021b. Disponível em: <https://help.altmetric.com/support/solutions/folders/6000237990>. Acesso em: 5 out. 2024.

BAAS, J. *et al.* Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, [s. l.], v. 1, n. 1, p. 377-386, 2020. DOI 10.1162/qss_a_00019. Disponível em: <https://direct.mit.edu/qss/article/1/1/377/15571/Scopus-as-a-curated-high-quality-bibliometric-data>. Acesso em: 5 out. 2024.

BALSTAD, M. T.; BERG, T. A long-term bibliometric analysis of journals influencing management accounting and control research. *Journal of Management Control*, [s. l.], v. 30, p. 357-380, 2020. DOI 10.1007/s00187-019-00287-8. Disponível em: <https://link.springer.com/article/10.1007/s00187-019-00287-8> Acesso em: 5 out. 2024.

BIRKLE, C. *et al.* Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, [s. l.], v. 1, n. 1, p. 363-376, 2020. DOI 10.1162/qss_a_00018. Disponível em: <https://direct.mit.edu/qss/article/1/1/363/15569/Web-of-Science-as-a-data-source-for-research-on>. Acesso em: 5 out. 2024.

BOSMAN, J.; KRAMER, B. Innovations in scholarly communication - data of the global 2015-2016 survey. *Zenodo*. [S. l.], 15 Apr. 2016. DOI 10.5281/zenodo.49583. Disponível em: <https://zenodo.org/records/49583> Acesso em: 5 out. 2024.

CABELLO VALDES, C. *et al.* *Evaluation of Research Careers fully acknowledging Open Science Practices: Rewards, incentives and/or recognition for researchers practicing Open Science: Working Group on Rewards under Open Science*. Luxembourg: European Commission, 2017. DOI 10.2777/75255. Disponível em: <https://euraxess.ec.europa.eu/sites/>

default/files/policy_library/os-rewards-wgreport-final_integrated_0.pdf.
Acesso em: 5 out. 2024.

CAGAN, R. The San Francisco Declaration on Research Assessment. *Disease Models & Mechanisms*, [s. l.], v. 6, n. 4, p. 869-870, 2013. DOI 10.1242/dmm.012955. Disponível em: <https://journals.biologists.com/dmm/article/6/4/869/3386/The-San-Francisco-Declaration-on-Research>
Acesso em: 5 out. 2024.

COSTAS, R. Towards the social media studies of science: social media metrics, present and future. *Bibliotecas*, La Habana, v. 13, n. 1, p. 1-5, enero./jun. 2017. Disponível em: <https://arxiv.org/pdf/1801.04437>. Acesso em: 5 out. 2024.

CROSSREF EVENT DATA. *Event Data User Guide*. [S. l.], 2021. Disponível em: <https://www.eventdata.crossref.org/guide/index.html>. Acesso em: 5 out. 2024.

EURITO CONSORTIUM. End of Pilot Phase Review. *Zenodo*. [S. l.], 2018. Disponível em: <https://zenodo.org/records/3413527>. Acesso em: 5 out. 2024.

EUROPEAN COMMISSION. *Open access to JRC Research Infrastructures: use the JRC Labs for Research*. [S. l.]: European Commision, 2020. Disponível em: <https://ec.europa.eu/jrc/en/research-facility/open-access>. Acesso em: 5 out. 2024.

EUROPEAN COMMISSION. *Open innovation, open science, open to the world: reflections of the Research, Innovation and Science Policy Experts (RISE) High Level Group*. Luxembourg: Publications Office of the European Union, 2018.

GARFIELD, E. The evolution of the science citation index. *International Microbiology*, [s. l.], v. 10, n. 1, p. 65-69, 2007. DOI 10.2436/20.1501.01.10. Disponível em: <https://garfield.library.upenn.edu/papers/barcelona2007a.pdf>. Acesso em: 5 out. 2024.

GARTNER. *Application programming interface (API)*. [S. l.], 2020. Disponível em: <https://www.gartner.com/en/information-technology/glossary/application-programming-interface>. Acesso em: 5 out. 2024.

GUÉDON, J.-C. *et al. Future of Scholarly Publishing and Scholarly Communication: Report of the Expert Group to the European Commission*.

Luxembourg: Publications Office of the European Union, 2019. DOI 10.2777/836532. Disponível em: <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1098&context=scholcom>. Acesso em: 5 out. 2024.

GUSENBAUER, M. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*, [Budapest], v. 118, p. 177-214, 2019. DOI 10.1007/s11192-018-2958-5. Disponível em: <https://link.springer.com/article/10.1007/s11192-018-2958-5>. Acesso em: 5 out. 2024.

HAUSCHKE, C.; CARTELLIERI, S.; HELLER, L. Reference implementation for open scientometric indicators (ROSI). *Research Ideas and Outcomes*, [s. l.], v. 4, 2018. DOI 10.3897/rio.4.e31656. Disponível em: <https://riojournal.com/article/31656/>. Acesso em: 5 out. 2024.

HEIBI, I.; PERONI, S.; SHOTTON, D. *Crowdsourcing open citations with CROCI: An analysis of the current status of open citations, and a proposal*. [S. l.], 2019. DOI 10.48550/arXiv.1902.02534. Disponível em: <https://arxiv.org/pdf/1902.02534>. Acesso em: 5 out. 2024.

HICKS, D. *et al.* (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature News*, [s. l.], v. 520, p. 429-431, 2015. Disponível em: <https://www.nature.com/articles/520429a>. Acesso em: 5 out. 2024.

LARIVIERE, V.; SUGIMOTO, C. R. The journal impact factor: A brief history, critique, and discussion of adverse effects. In: GLÄNZEL, W. *et al.* *Springer handbook of science and technology indicators*. [S. l.]: Springer, 2019. p. 3-24.

LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY (Alemanha). About this Registry. In: LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY (Alemanha). *Registry of Scientometric Data Sources*. Alemanha, [20--a]. Disponível em: <https://osl.tib.eu/rosi-registry/about.php>. Acesso em: 20 ago. 2023.

LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY (Alemanha). COCI. In: LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY (Alemanha). *Registry of Scientometric Data Sources*. Alemanha, [20--c]. Disponível em: <https://osl.tib.eu/rosi-registry/display.php?sourceId=12>. Acesso em: 20 ago. 2023.

LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY (Alemanha). Home. In: LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY (Alemanha). *Registry of Scientometric Data Sources*. Alemanha, [20--b]. Disponível em: <https://osl.tib.eu/rosi-registry/index.php>. Acesso em: 20 ago. 2023.

LEIBNIZ INFORMATION CENTRE FOR SCIENCE AND TECHNOLOGY (Alemanha). *Registry of Scientometric Data Sources*. Alemanha, [20--d]. Disponível em: <https://osl.tib.eu/rosi-registry/>. Acesso em: 20 ago. 2023.

LEMKE, S. *et al.* "When You Use Social Media You Are Not Working": Barriers for the Use of Metrics in Social Sciences. *Frontiers in Research Metrics and Analytics*, [s. l.], v. 3, n. 39, p. 1, Jan. 2019. DOI 10.3389/frma.2018.00039. Disponível em: <https://www.frontiersin.org/journals/research-metrics-and-analytics/articles/10.3389/frma.2018.00039/full>. Acesso em: 5 out. 2024.

LILIENTHAL, S. Registry of [Open] Scientometric Data Sources: a collaborative directory of scientometric data sources. *TIB Blog*, [s. l.], 2019. Disponível em: <https://blogs.tib.eu/wp/tib/2019/05/23/registry-of-open-scientometric-data-sources-a-collaborative-directory-of-scientometric-data-sources/>. Acesso em: 5 out. 2024.

LOTKA, A. J. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, [s. l.], v. 16, n. 12, p. 317-323, June 1926. Disponível em: https://www.jstor.org/stable/pdf/24529203.pdf?refreqid=fastly-default%3Ab41e14d360c92b44f234b0b28eefb98c&ab_segments=&initiator=&acceptTC=1. Acesso em: 5 out. 2024.

MANGHI, P.; ATZORI, C. The Data4Impact Knowledge Base Platform (D4IP). *Platform high-level architecture and data flows*. [S. l.], 2018. Disponível em: <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5be308c39&appId=PPGMS>. Acesso em: 5 out. 2024.

MOHER, D. *et al.* The Hong Kong Principles for assessing researchers: Fostering research integrity. *PLOS Biology*, [s. l.], v. 18, n. 7, p. 1-14, 2020. DOI 10.1371/journal.pbio.3000737. Disponível em: <https://journals.plos.org/plosbiology/article/file?id=10.1371/journal.pbio.3000737&type=printable>. Acesso em: 5 out. 2024.

MUGNAINI, R. *et al.* Openness trends in Brazilian citation data: factors related to the use of DOIs. *Scientometrics*, [Budapest], v. 126, n. 3, p. 2523-2556, 2021. DOI 10.1007/s11192-020-03663-7. Disponível em: <https://link.springer.com/article/10.1007/s11192-020-03663-7>. Acesso em: 5 out. 2024.

NATIONAL INFORMATION STANDARDS ORGANIZATION. *NISO Alternative Assessment Metrics (Altmetrics) Initiative*. Baltimore: NISO, 2016. Disponível em: <https://www.niso.org/standards-committees/altmetrics>. Acesso em: 5 out. 2024.

OPEN SCIENCE MONITOR. *Updated methodological note*. Brussels: OSM, 2019. Disponível em: https://research-and-innovation.ec.europa.eu/system/files/2020-01/open_science_monitor_methodological_note_april_2019.pdf. Acesso em: 5 out. 2024.

PHILANDIA. Ministry of Education and Culture. *Evaluation of Openness in the Activities of Research Organisations and Research Funding Organisations in 2016*. Philandia: Ministry of Education and Culture, 2016. Disponível em: https://www.doria.fi/bitstream/handle/10024/127273/Evaluation_of_Openness_2016_FIN.pdf?sequence=2&isAllowed=y. Acesso em: 5 out. 2024.

PID SERVICES REGISTRY. *About*. [S. l.], 2020. Disponível em: <https://pidservices.org/about>. Acesso em: 5 out. 2024.

PLUM ANALYTICS. *About PlumX Metrics*. [S. l.], 2019. Disponível em: <https://plumanalytics.com/learn/about-metrics/>. Acesso em: 5 out. 2024.

PRIEM, J. *et al.* *Altmetrics: a manifesto*. [S. l.], 26 Oct. 2010. Disponível em: <http://altmetrics.org/manifesto/>. Acesso em: 5 out. 2024.

SÏLE, L. *et al.* *European Databases and Repositories for Social Sciences and Humanities Research Output*. Antwerp: ECOOM & ENRESSH, 2017. DOI 10.6084/m9.figshare.5172322. Disponível em: https://enressh.eu/wp-content/uploads/2017/09/2017_ENRESSH_European_Databases.pdf. Acesso em: 5 out. 2024.

STUART, D. Open bibliometrics and undiscovered public knowledge. *Online Information Review*, [Brandford], v. 42, n. 3, p. 412-418, 2018.

SUGIMOTO, C. R.; MURRAY, D. S.; LARIVIÈRE, V. Open citations to open science. *ISSI*. [S. l.], 26 Apr. 2018. Disponível em: <https://www.issi-society.org/blog/posts/2018/april/open-citations-to-open-science/>. Acesso em: 5 out. 2024.

TAY, A.; KRAMER, B.; WALTMAN, L. Why openly available abstracts are important - overview of the current state of affairs. *Leiden Madtrics*, [s. l.], 2020. Disponível em: <https://leidenmadtrics.nl/articles/why-openly-available-abstracts-are-important-overview-of-the-current-state-of-affairs>. Acesso em: 5 out. 2024.

THELWALL, M.; VAUGHAN, L.; BJÖRNEBORN, L. Webometrics. *Annual Review of Information Science and Technology*, White Plains, v. 39, n. 1, p. 81-135, 2005.

TIPPET, C.; BLIND, K.; PARRAGUEZ RUIZ, P. Audit of new R&I data. *Zenodo*. [S. l.], 2018. Disponível em: <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5bb228548&appId=PPGMS>. Acesso em: 5 out. 2024.

VISSER, M.; VAN ECK, N. J.; WALTMAN, L. Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, [s. l.], v. 2, n. 1, p. 20-41, 2021. DOI 10.1162/qss_a_00112. Disponível em: <https://direct.mit.edu/qss/article/2/1/20/97574/Large-scale-comparison-of-bibliographic-data>. Acesso em: 5 out. 2024.

VON SCHOMBERG, R. (ed.). *Indicator frameworks for fostering open knowledge practices in science and scholarship*: Report of the Expert Group on Indicators for Researchers' engagement with open science. Brussels: European Commission, 2019.

WALTMAN, L.; LARIVIÈRE, V. Special issue on bibliographic data sources. *Quantitative Science Studies*, [s. l.], v. 1, n. 1, p. 360-362, 2020. DOI 10.1162/qss_e_00026. Disponível em: <https://direct.mit.edu/qss/article/1/1/360/15575/Special-issue-on-bibliographic-data-sources>. Acesso em: 5 out. 2024.

WALTMAN, L. *et al.* Opening science: The rebirth of a scholarly journal. *Quantitative Science Studies*, [s. l.], v. 1, n. 1, p. 1-3, 2020. DOI 10.1162/qss_e_00025. Disponível em: https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/23210/qss_e_00025.pdf?sequence=1&isAllowed=y. Acesso em: 5 out. 2024.

WASS, J. Five principles for community altmetrics data. altmetrics18. *Joé's Blog*, London, 2018. Disponível em: <https://blog.afandian.com/2018/05/five-principles-altmetrics/>. Acesso em: 5 out. 2024.

WILSDON, J. *et al.* *Next-generation metrics*: Responsible metrics and evaluation for open science: Report of the European Commission Expert Group on Altmetrics. Brussels: European Commission, 2017.

ZAHEDI, Z.; COSTAS, R. General discussion of data quality challenges in social media metrics: Extensive comparison of four major altmetric data aggregators. *PLoS ONE*, [s. l.], v. 13, n. 5, p. 1-27, 2018. DOI 10.1371/journal.pone.0197326. Disponível em: <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0197326&type=printable>. Acesso em: 5 out. 2024.

CAPÍTULO 9

Metadados na comunicação científica: mapeamento de modelos e aplicações

Fernando Modesto

Amanda Ramalho

Introdução

A comunicação científica pode ser entendida como o processo de troca de informações entre os membros de uma comunidade especializada, com o objetivo de divulgar os resultados de pesquisas realizadas conforme normas estabelecidas e controladas pelo contexto acadêmico. Esse intercâmbio é essencial para o desenvolvimento do conhecimento, permitindo que descobertas e teorias sejam compartilhadas, avaliadas e validadas por outros pesquisadores (Barbalho, 2005). Assim, configura-se como um conceito chave para expressar a qualidade e a dinâmica do desenvolvimento e da divulgação do conhecimento gerado por esta determinada comunidade. Além de promover e dar visibilidade aos saberes produzidos, permite a análise, comprovação e medição das pesquisas científicas realizadas.

As cinco décadas, desde que Diana Crane escreveu seu artigo sobre a natureza da comunicação e a influência científica, testemunharam uma reestruturação radical da comunicação acadêmica. Na oportunidade, ela observou que o enorme crescimento de novos conhecimentos está exigindo maior flexibilidade no sistema de comunicação formal e que o progresso no uso desse sistema exigirá maior compreensão sobre como os cientistas usam ideias e sobre as próprias ideias (Crane, 1972). Os modelos foram atualizados para refletir as mudanças provocadas pelas tecnologias de informação emergentes, particularmente em campos como ciência espacial, física de alta energia e pesquisa do genoma humano. Talvez o resultado mais importante da incursão da tecnologia da informação no trabalho científico não tenha

sido a velocidade, a flexibilidade ou o alcance da comunicação mediada por computador, mas as consequências que esses atributos tiveram para a prática da ciência. Novas questões fundamentais estão sendo levantadas sobre a produção, a transferência e o acesso a informações acadêmicas que não eram feitas antes da revolução eletrônica (Russell, 2001).

A ciência, por meio de sua produção e comunicação científica, é impactada pela evolução contínua do ambiente digital. Esta tem mudado, continuamente, a maneira de tratar, organizar, armazenar, acessar, comunicar, compartilhar e reutilizar as informações e conhecimentos científicos. Outro aspecto é que esse cenário tem possibilitado às comunidades de conhecimento a publicação de seus trabalhos acadêmicos em diversos formatos, além do domínio do texto. Portanto, recursos tecnológicos e digitais constituem alternativas promissoras para a colaboração e compartilhamento da produção científica. Produção que flui por repositórios de *preprints*, de revistas científicas, livros e capítulos de livros, de teses e dissertações, principalmente, de acesso aberto. Assim, o seu controle, tratamento e organização impõem adoção de padrões de metadados que permitam torná-la acessível e utilizável.

Os padrões de metadados são ferramentas essenciais na organização de dados em contextos específicos. O uso de metadados na descrição de objetos ajuda a reduzir as dificuldades associadas à localização, descoberta, documentação, seleção e recuperação da informação (Baptista; Ferneda, 2016).

Metadados são os dados estruturados sobre as fontes de informação. Uma chave para recuperar informações que forneçam pontos de acesso significativos para os pesquisadores. Existem variados esquemas e padrões de metadados que são mais adequados às diferentes comunidades científicas e profissionais, de acordo com suas necessidades e propósitos.

Uma abordagem sobre a questão envolve analisar “se os metadados presentes na comunicação científica, podem ser aplicados indistintamente ao tratamento e organização das informações atinentes?”. Também, no contexto da abordagem, identificar quais formatos de metadados são mais utilizados no âmbito da comunicação.

Nesse sentido, este texto procura destacar a importância dos metadados no processo da comunicação científica. Fornece uma análise geral do conceito e da tipologia de metadados e descreve alguns padrões, que exemplificam

a evolução das suas aplicações no campo da comunicação científica. Toma como suporte de análise uma breve revisão bibliográfica sobre o tema.

Comunicação científica: definição

A comunicação científica é o processo pelo qual a pesquisa é conduzida, transformada em conteúdo e distribuída para um público mais amplo (Gregg *et al.*, 2019). Assim, as informações científicas consistem em conteúdos e materiais gerados por e para acadêmicos – estudantes, pesquisadores, cientistas, professores e outros. A finalidade é aprender ou avançar no conhecimento sobre tópicos científicos específicos. Os recursos são produzidos para compartilhar e comunicar pesquisas. Normalmente, assume a forma de textos, apresentações, artigos de periódicos, livros e capítulos de livros etc. As publicações acadêmicas incluem conteúdo pesquisado, referências citadas – por exemplo, notas de rodapé e citações no texto – e bibliografias. Esses elementos facilitam o compartilhamento de conhecimento, listam outras fontes para acompanhamento e demonstram que novas ou antigas ideias se baseiam em pesquisas anteriores. Outro aspecto único, da informação acadêmica, refere-se à revisão por pares. Um padrão de revisão editorial que ocorre antes que um texto acadêmico possa ser publicado. Os revisores são especialistas no assunto tratado pelo conteúdo avaliado. Suas leituras e revisões críticas ajudam a verificar se a pesquisa é válida e se as conclusões do pesquisador não são apenas sólidas, mas também importantes. As bibliotecas acadêmicas coletam esses recursos revisados por pares, com um enfoque nas áreas de assunto ensinadas em sua instituição (Iowa State University, 2019).

A comunicação científica é essencial para a natureza da ciência. O meio de aferir o número de publicações e artigos indica o crescimento científico e a validação do processo de pesquisa que gera essa produção. Os pesquisadores não só comunicam os resultados aos seus pares, mas também à sociedade em geral, por meio dos canais de comunicação. Eles também se apoiam no conhecimento registrado nos canais de comunicação. O intercâmbio de opiniões e de dados entre pares é um aspecto importante da produção científica (Russell, 2001).

Nesse sentido, a comunicação faz-se presente nas etapas do processo de pesquisa e na sua posterior divulgação, bem como da sua posterior recuperação para subsidiar outras investigações. Essa recuperação se realiza pela utilização de um sistema de informação baseado em estrutura de dados ou de metadados. Dessa forma, a comunicação científica pode ser entendida como a infraestrutura informacional da comunidade científica (Lyman, 1997).

Para entender a comunicação científica e os metadados, há que considerar o contexto em que essa comunicação opera. Como destaca Mueller (2006), da mesma maneira que existe hierarquia entre os indivíduos que formam as comunidades científicas, há uma hierarquia entre os diversos tipos de veículos usados para comunicar o texto científico – por exemplo: periódico, livro, trabalho de eventos. Apesar de variar, conforme a área de conhecimento, o periódico indexado detém prestígio perante os sistemas de avaliação da pesquisa e de programas de pós-graduação, baseados em métricas de citação. Entretanto, entre estes, também há publicações que atribuem maiores graus de prestígio às editoras responsáveis. Desde o século XIX, a comunicação científica centra-se em torno dos periódicos, como canal de divulgação para a produção da maior parte dos trabalhos científicos avaliados (Fredriksson, 2001).

Greig (2019) comenta que os metadados no contexto da comunicação científica acompanham as várias etapas e resultados da pesquisa. Nesse sentido, os elementos de metadados são comuns à maioria das pesquisas acadêmicas, como autor, data, título, assunto, idioma e identificador padrão. No caso de dados de pesquisa, os metadados descrevem aspectos especializados, como a localização geográfica onde os dados foram coletados, o nome e o identificador do financiador da pesquisa, a afiliação institucional dos pesquisadores e colaboradores, como editores e curadores de dados.

Entretanto, a revolução digital desconstruiu e fragmentou a unidade do texto, erodindo a noção do “documento” monolítico. Ao mesmo tempo – e, em partes, devido à fragmentação –, autores e editores enriquecem o artigo com novos conteúdos e formatos de recursos, como multimídia. Na atualidade, os resultados científicos são divulgados como conjuntos de dados, em formatos digitais (Schöpfel, 2014). Assim, torna-se ambiente propício à

valorização dos metadados como meio de tornarem os documentos científicos ativos. Eles se tornam janelas para os resultados de uma pesquisa.

Metadados

Esta seção apresenta os conceitos e definições de metadados, como recursos que descrevem outros dados e facilitam o acesso à informação. São abordadas três tipologias principais de metadados: os descritivos, que identificam e localizam recursos; os administrativos, relacionados à gestão e preservação; e os estruturais, que indicam a organização interna dos recursos. Essas categorias são essenciais para a eficiência na recuperação da informação.

Conceitos e definições

Um dos primeiros periódicos impressos a dedicar número especial ao tema dos metadados e à dimensão de sua abrangência e aplicação foi a revista *Bulletin of the American Society for Information Science* (Bulletin, 1997), que reuniu diversos artigos relacionados à temática. Todavia, a origem do termo aparece na dissertação de doutorado de Philip R. Bagley, e como um relatório apresentado ao Air Force Office of Scientific Research, intitulado *Extension of Programming Language Concepts* – Relatório Técnico do University City Science Center, novembro de 1968 (Bagley, 1969, 2014).

A definição clássica de metadados é literal e baseada na própria etimologia da palavra, são “dados sobre dados”. Segundo Pomerantz (2015, p. 30), essa definição não contempla a magnitude das funções que o termo exerce. Os dados podem ser apenas informações potenciais, brutas e não processadas, conseqüentemente, também sem sentido. Entretanto, se a palavra “dados” for compreendida como “informação potencial”, poderá ser entendida não como um conceito abstrato, mas como objeto potencialmente informativo. Nesse sentido, os metadados passam a ser definidos como “objeto potencialmente informativo para descrever outro objeto potencialmente informativo”.

A National Information Standard Organization (Niso) conceitua metadados como “a informação estruturada que descreve, explica, localiza, ou

possibilita que um recurso informacional seja fácil de recuperar, usar ou gerenciar” (National Information Standards Organization, 2004, p. 1).

Entretanto, a distinção fundamental entre metadados e conteúdo é que os dados são o próprio conteúdo que fornece uma descrição, medição ou até mesmo um relatório sobre qualquer coisa relacionada aos ativos de dados de um recurso ou de uma organização. Por outro lado, os metadados descrevem as informações relevantes sobre esses dados, dando-lhes mais contexto para os usuários. Enquanto os dados podem ser processados ou não, como dados brutos – números ou caracteres não informativos –, os metadados são sempre considerados informações processadas. Alguns dados podem ser informativos, enquanto outros podem não ser. No entanto, os metadados são sempre informativos, pois fazem referência a outros dados (Metadata [...], 2023).

Em realidade, a comunidade científica desenvolve sistemas de informação que se baseiam na arquitetura de rede, como a internet, e explicita uma estruturação de informação e um padrão de metadados. Entretanto, os sistemas bibliográficos tradicionais denominam seus registros bibliográficos, como “metadados”. Os sistemas de informação científica têm adotado, na descrição de seus recursos, estrutura semelhante aos sistemas bibliográficos, ou seja, estruturas de dados que organizam as informações científicas ou acadêmicas. Independente do contexto, o termo “metadados” se tornou recorrente por diversas comunidades das áreas científicas, com diferentes finalidades e motivações. No entanto, todas essas áreas compartilham, na maioria dos casos, ferramentas, padrões e questionamentos em torno dos metadados (Méndez Rodríguez, 2001).

Seguindo essa linha, um aspecto relevante na definição dos metadados é o nível de sua aplicação nos diversos contextos sociais, científicos e de interoperabilidade em rede. Os metadados se aplicam em muitos domínios de distintas áreas de conhecimento que requeiram tratamentos específicos de dados, como, por exemplo: agências de fomento, órgãos públicos, organizações do patrimônio cultural, instituições de ensino e pesquisa e bases de dados em geral. Na comunicação científica, os metadados tornam-se importantes para o tratamento de livros, revistas, periódicos, textos de eventos, relatórios e dados de pesquisa, favorecendo a distinção entre os diferentes

objetos, a normalização, a interoperabilidade e a recuperação de seu conteúdo.

Assim, metadados são estruturas descritivas e classificatórias sobre objetos ou recursos digitais. Em sistemas computacionais, eles são aplicados como fonte de informação para identificar e promover a uniformização do conteúdo proveniente de sistemas externos, por meio do intercâmbio de dados. A área da biblioteconomia utiliza do metadado para melhorar a capacidade de armazenamento informacional, com reflexo na melhoria do sistema de recuperação da informação e comunicação (Souza, 2021).

Dessa forma, o compartilhamento de informação proporciona agilidade em atividades operacionais e de recuperação da informação, desde a inserção e o registro até o acesso dos elementos informacionais, de acordo com a disponibilidade do recurso. Nesse contexto, os metadados juntamente com os dados desempenham um papel fundamental no compartilhamento. As comunidades e os provedores de conteúdo relacionados à produção, à divulgação, ao armazenamento ou à preservação de conteúdo estão em constante busca por refinamentos e acordos na semântica de metadados especializados para permitir novos formatos internacionais (Ramesh; Vivekavardhan; Bharathi, 2015).

Certamente, não há um tipo único de esquema de metadados que atenda a todos os aplicativos, tipos de recursos e de comunidades de informação. Segundo Ramesh, Vivekavardhan e Bharathi (2015), uma estrutura de metadados pode apresentar cinco componentes principais: (1) esquema-padrão com elementos e categorias de informação a serem escolhidas para registro; (2) vocabulário específico e seus valores escolhidos para as categorias definidas; (3) estrutura de modelo conceitual que descreve como diferentes conceitos informações estão relacionadas umas com as outras; (4) padrão de conteúdo, referente à forma de registrar as informações específicas – por exemplo, vocabulários –, que são inseridas nas categorias esquemática de metadados; e (5) codificação – orientada com a maneira como os metadados serão apresentados, por exemplo, XML.

Para BasuMallick (2022), as funções dos metadados devem servir:

- organização, descrição e ordenamento dos recursos de dados em um repositório. As informações são organizadas com base no público ou tópico. Ademais, pela construção dinâmica de páginas usando informações armazenadas no banco de dados. Os especialistas podem produzir novos metadados por meio de procedimentos de registro, catalogação e indexação;
- pesquisa e recuperação que facilitem localizar e obter dos metadados informações de recursos de interesse dos usuários;
- utilização e preservação que ajudem a rastrear o ciclo de vida de um recurso;
- criação de informações e versões, além da reutilização de um recurso. Deve-se incluir metadados administrativos e descritivos sobre o recurso;
- facilitação da interoperabilidade pelo uso de esquemas de metadados, mecanismos de transferência padronizados e cruzamentos entre esquemas e Interfaces de Programação de Aplicativos (APIs) para compartilhar recursos, como coleta de metadados e pesquisa entre sistemas.

A criação de metadados decorre da elaboração de registros essenciais à identificação e recuperação de um recurso informacional ou bibliográfico. Sendo assim, sempre existiram metadados para identificar características de um recurso, no entanto, a revolução e evolução digitais aumentaram o seu valor. Ademais, a atividade da catalogação bibliográfica caracterizou-se pela sua tradição e finalidade em uma ação propícia à criação de metadados. Entretanto, enquanto a catalogação tradicional resulta na produção de catálogo bibliográfico, segundo critérios definidos por normas catalográficas, a elaboração de metadados introduz esquemas para codificar dados, à semelhança da representação dos recursos bibliográficos (Guerrini, 2022).

Metadados: tipologias

Nas últimas décadas, os metadados têm sido reconhecidos como um importante componente de realização e resultado de pesquisa. Essa tendência é

definida por inúmeras iniciativas de desenvolvimento de padrões variados de metadados. Nesse desenvolvimento, há um consenso em relação aos três tipos de padrões de metadados divididos em categorias, conforme seu uso e objetivo específicos, sendo estes: administrativos/técnicos, estruturais e descritivos (Riley, 2017; Sayão, 2010).

Metadados administrativos

Os metadados administrativos englobam uma ampla gama de informações sob um “guarda-chuva” que apoia aspectos técnicos, de gestão e de operação relacionados ao uso de recursos. Nesse ínterim, eles incluem metadados técnicos, de preservação e de direitos, que são tratados pela sua sobreposição de usos. Os metadados de preservação, por exemplo, fornecem informações de apoio aos processos que garantem ao recurso continuidade de existência ao longo do tempo. Os metadados de direitos fornecem informações usadas para controlar o acesso ao recurso, sob condições legais de seu uso. Certamente, tal cuidado é uma forma de administração (Leipzig *et al.*, 2021; Pomerantz, 2015).

Assim, os metadados administrativos fornecem dados referentes à gestão, organização, disponibilização e modo de acesso aos recursos e/ou acervos digitais (Ramesh; Vivekavardhan; Bharathi, 2015). O Quadro 1 apresenta exemplos de metadados desta categoria.

Quadro 1 – Metadados administrativos aplicáveis à comunicação científica

Nº	Metadados	Detalhe e Fonte
1	ANSI/Niso Z39.87	<i>Niso Technical Metadata for Digital Still Images</i> . Padrão de imagens digitais (TIFF, JPEG, GIF). Baseado em XML. Mantido pela Library of Congress e pelo Niso Technical Metadata for Digital Still Images Standards Committee ¹ .
2	CDE	<i>Common Data Elements</i> . Esquema de paradados clínicos e de pesquisa na área da saúde ² .
3	CSDGM	<i>Content Standard for Digital Geospatial Metadata</i> . Padrão para mapas, de autoria do Comitê Federal de Dados Geográficos (FGDC) ³ .

1 Disponível em: <https://bit.ly/3AsLMPK>.

2 Disponível em: <https://cde.nlm.nih.gov/home>.

3 Disponível em: <https://www.fgdc.gov/metadata/csdgm-standard>.

Nº	Metadados	Detalhe e Fonte
4	EXIF	<i>Exchangeable Image File Format</i> . Padrão de imagem de câmera fotográfica – data, hora, localização. Usado para autenticar foto ⁴ .
5	FITS	<i>Flexible Image Transport System</i> . Formato para armazenar e manipular imagens e dados astronômicos ⁵ .
6	IPTC	<i>International Press Telecommunications Council Photo Metadata Standard</i> . Padrão descritivo de imagens para agências de notícias, fotógrafos, bibliotecas, museus e empresas relacionadas ⁶ .
07	ISO 19115-1: 2014	<i>ISO 19115-1:2014 Geographic information — Metadata</i> . Esquema para mapas, gráficos e documentos textuais ⁷ .
08	METS	<i>Metadata Encoding and Transmission Standard</i> . Padrão descritivo, administrativo e estrutural para descrever objetos digitais, usando linguagem XML. Mantido pelo MARC Standards Office da Library of Congress ⁸ .
09	MIME	<i>Multipurpose Internet Mail Extensions</i> . Extensão do protocolo Simple Mail Transport Protocol (SMTP). Para troca de arquivos – áudio, vídeo, imagens ⁹ .
10	MIX	<i>Niso Metadata for Images in XML</i> (ANSI/NISO 39.87-2006). Metadados técnicos para gestão de imagens, baseado em XML. Mantido pela Library of Congress ¹⁰ .
11	MPEG-7	<i>Moving Pictures Expert Group</i> . Metadados descritivos para imagem e áudio digital. Baseado em XML ¹¹ .
12	PREMIS	<i>PREservation Metadata Implementation Strategies</i> . Padrão para preservação de objetos digitais. Mantido pela Library of Congress ¹² .
13	VideoMD AudioMD	Metadados para objetos digitais em áudio e vídeo. Extensão para METS e PREMIS ¹³ .
14	XMP	<i>Extensible Metadata Platform</i> . Esquema para informações de direitos autorais, palavras-chave e outros dados, em arquivos digitais, como imagens, áudio e vídeo. Mantido pela Adobe Systems ¹⁴ .

Fonte: elaborado pelos autores.

4 Disponível em: <https://adobe.ly/3NcQhG0>.

5 Disponível em: <https://bit.ly/41L1hP8>.

6 Disponível em: <https://iptc.org/standards/photo-metadata/iptc-standard/>.

7 Disponível em: <https://www.iso.org/standard/53798.html>.

8 Disponível em: <https://bit.ly/3V68kzC>.

9 Disponível em: <https://en.wikipedia.org/wiki/MIME>.

10 Disponível em: <https://www.loc.gov/standards/mix/>.

11 Disponível em: <https://mpeg.chiariglione.org/>.

12 Disponível em: <https://www.loc.gov/standards/premis/>.

13 Disponível em: <https://www.loc.gov/standards/amdvmd/>.

14 Disponível em: <https://www.adobe.com/products/xmp.html>.

Metadados estruturais

Os metadados estruturais são essenciais para identificar a estrutura lógica de um objeto, e das suas relações internas e externas, como, por exemplo, entre as páginas de um documento, capítulos de um livro, ou qualquer outro objetivo que possa ser organizado hierarquicamente. No ambiente digital, os recursos lógicos podem ser baseados em um conjunto de vários arquivos físicos. Os metadados estruturais relacionam esses arquivos entre si e com a estrutura dos objetos lógicos. Por exemplo, um livro com 100 páginas pode ter sido digitalizado como 100 imagens em formato Tag Image File Format (TIFF). Cada imagem representa uma página. Esses metadados são necessários para indicar qual arquivo TIFF é a página buscada, qual a ordem correta de cada página e, também, para indicar quais páginas compõem o capítulo 1 ou o capítulo 2, por exemplo. Com essas informações, um aplicativo pode ser desenvolvido e executado para exibir o conteúdo de forma organizada, como sumário ou capítulos para o leitor.

O Quadro 2 apresenta exemplos que se enquadram na categoria de metadados estruturais. Estes podem associar diferentes representações de um mesmo conteúdo intelectual. Além disso, em um contexto diferente, podem estar relacionados ao gerenciamento, à preservação e à exibição dos recursos. No caso de um repositório, com a responsabilidade de armazenar recursos digitais, é fundamental saber das características dos arquivos para sua correta exibição. Em alguns contextos, os sistemas de marcação, como SGML ou XML, sob diretrizes TEI, inclui informações estruturais, como quebras de página, cabeçalhos de capítulos e outras divisões (Leipzig *et al.*, 2021; Ramesh; Vivekavardhan; Bharathi, 2015).

Quadro 2 – Metadados estruturais aplicáveis à comunicação científica

Nº	Metadados	Detalhe e Fonte
15	BITS	<i>Book Interchange Tag Set</i> . Para descrever o conteúdo estrutural e semântico de livros e seus componentes. Baseado em XML e JATS ¹⁵ .
16	CIDOC/CRM	Conceptual Reference Model (CRM). Ontologia para informações do patrimônio cultural, galerias, bibliotecas, arquivos, museus. Mantido pela International Committee for Documentation of the International Council of Museums (ICOM) ¹⁶ .

15 Disponível em: <https://jats.nlm.nih.gov/extensions/bits/>.

16 Disponível em: <https://cidoc-crm.org/>.

Nº	Metadados	Detalhe e Fonte
17	CrossRef	<i>CrossRef Metadata Schema</i> . Pesquisar metadados de artigos de periódicos, anais de eventos, preprints, dissertações e livros ¹⁷ .
18	DataCite	<i>DataCite Metadata Schema</i> . Identificação de recurso para fins de citação e recuperação ¹⁸ .
19	DDI	<i>Data Documentation Initiative</i> . Para dados de pesquisas e métodos observacionais nas ciências sociais, comportamentais, econômicas e da saúde ¹⁹ .
20	DocBook	Esquema para produção de livros e artigos em vários formatos impressos e <i>on-line</i> . Baseado em SGML e XML. Mantido pelo OASIS ²⁰ .
21	e-GMS	<i>e-Government Metadata Standard</i> . Padrão do governo eletrônico do Reino Unido. Para rotular o conteúdo. Baseado no <i>Dublin Core Metadata Element Set</i> ²¹ .
22	EML	<i>Ecological Metadata Language</i> . Especificação em XML para documentar, preservar e compartilhar dados de pesquisa nas ciências da terra e meio ambiente ²² .
23	FRBRoo	<i>Functional Requirements for Bibliographic Records-object oriented</i> . Ontologia para informação bibliográfica e o intercâmbio com a informação museológica, por meio do modelo CIDOC CRM ²³ .
24	HTML	<i>Hyper-Text Markup Language</i> . Mantido pelo World Wide Web Consortium ²⁴ .
25	JATS	<i>Journal Article Tag Suite</i> : NSI/NISO Z39.96-2015. Modelar a estrutura e a semântica dos artigos de periódicos, baseado em XML ²⁵ .
26	LOM	<i>IEEE Learning Object Metadata</i> . Modelo do <i>Institute of Electrical and Electronics Engineers</i> , baseado em XML. Descreve objetos de aprendizagem digital ²⁶ .
27	LRMoo	<i>Library Reference Model object-oriented</i> . Ontologia para representação bibliográfica e intercâmbio com a museológica. Mantido pelo IFLA LRMoo Working Group e CIDOC CRM Special Interest Group ²⁷ .
28	MeSH	<i>Medical Subject Headings</i> . Vocabulário para publicações da área da saúde. Mantido pela National Library of Medicine (NLM) ²⁸ .

17 Disponível em: <https://www.crossref.org/documentation/schema-library/>.

18 Disponível em: <https://schema.datacite.org/>.

19 Disponível em: <https://ddialliance.org/>.

20 Disponível em: <https://docbook.org/whatis>.

21 Disponível em: <https://en.wikipedia.org/wiki/E-GMS>.

22 Disponível em: <https://eml.ecoinformatics.org/>.

23 Disponível em: http://cidoc.ics.forth.gr/frbr_inro.html.

24 Disponível em: <https://pt.wikipedia.org/wiki/HTML>.

25 Disponível em: <https://jats.nlm.nih.gov/>.

26 Disponível em: https://en.wikipedia.org/wiki/Learning_object_metadata.

27 Disponível em: <https://bit.ly/43Wve05>.

28 Disponível em: <https://www.nlm.nih.gov/mesh/meshhome.html>.

Nº	Metadados	Detalhe e Fonte
29	MIDAS Heritage	<i>MIDAS Heritage: the UK Historic Environment Data Standard</i> . Para registro de ativos digitais britânicos de patrimônio histórico e cultural. Mantido pela English Heritage ²⁹ .
30	EPUB	Formato que representa, empacota e codifica conteúdos de forma estruturada e semanticamente. Inclui HTML, CSS, XML, entre outros. Mantido pelo W3C ³⁰ .
31	ONIX for Books	<i>ONLine Information eXchange for Books</i> . Padrão baseado em XML, para livros impressos e eletrônicos. Mantido pelo Digital Issues working group of the Association of American Publishers e EDItEUR ³¹ .
32	ORCID	<i>Open Researcher and Contributor ID</i> . Identificador para distinguir pesquisadores com ambiguidade de nomes ³² .
33	PBCore	<i>Public Broadcasting Metadata Dictionary</i> . Padrão para conteúdo audiovisual. Baseado em XML. Desenvolvido pela comunidade de radiodifusão pública norte-americana ³³ .
34	RDF	<i>Resource Description Framework</i> . Para intercâmbio de dados na web. Mantido pelo W3C ³⁴ .
35	Schema.org	Esquema para visibilidade de conteúdos em buscadores. Mantido pelo W3C ³⁵ .
36	SGML	<i>Standard Generalized Mark-up Language</i> . Padrão para marcação de documentos. Norma ISO 8879:1986 ³⁶ .
37	SMPTE	<i>Society of Motion Picture Technical Experts Data Dictionary</i> . Desenvolvedora de padrões para imagens médicas, vídeo, filme e áudio digital ³⁷ .
38	SPECTRUM	<i>The UK Museum Management Standard</i> . Norma para gestão de coleções de museus. Mantida por entidades museológicas ³⁸ .
39	TEI	<i>Text Encoding Initiative</i> . Padrão, baseado em SGML, para marcação de textos ³⁹ .
40	XML	<i>Extensible Mark-up Language</i> . Para intercâmbio de dados na web. Mantido pelo W3C ⁴⁰ .

Fonte: elaborado pelos autores.

29 Disponível em: <http://www.heritage-standards.org.uk/midas-heritage/>.

30 Disponível em: <https://www.w3.org/TR/epub-33/>.

31 Disponível em: <https://www.editeur.org/83/Overview/>.

32 Disponível em: <https://orcid.org/>.

33 Disponível em: <https://pbcore.org/>.

34 Disponível em: <https://www.w3.org/RDF/>.

35 Disponível em: <https://schema.org/docs/about.html>.

36 Disponível em: <https://bit.ly/3VKArla>.

37 Disponível em: <https://www.smpte.org/>.

38 Disponível em: <https://collectiontrust.org.uk/spectrum/spectrum-5/>.

39 Disponível em: <https://tei-c.org/>.

40 Disponível em: <https://www.w3.org/XML/>.

Metadados descritivos

Os metadados descritivos apoiam a descoberta e a avaliação geral de um recurso, fornecendo informações detalhadas sobre as características, formato, conteúdo e atributos de um recurso digital. Na perspectiva científica, padrões descritivos permitem a documentação, descoberta e interoperabilidade da pesquisa científica. Juntamente com os metadados administrativos, também fornecem contexto e histórico sobre a origem, a autenticidade e o ciclo de vida dos dados. Esses padrões podem ser incorporados na produção científica de tabelas e listas, formatadas em diversos tipos de arquivos e banco de dados, servindo como fonte para análises computacionais reproduzíveis, filtrando visualizações e estatísticas em pesquisas científicas (Leipzig *et al.*, 2021).

O Quadro 3 apresenta exemplos de metadados que se enquadram na categoria de descritivos. Fornecem informações de curadoria, como nome do arquivo, criador, datas associadas, descrição, resumo, localizações etc., sendo uma fonte para que usuários possam encontrar, acessar e identificar o recurso. Esses metadados lidam com dados relacionados ao contexto de um recurso de informação e à sua criação.

Existem dois modelos de metadados descritivos comumente citados: (I) modelo de metadados separado, no qual os metadados descritivos são armazenados em um banco de dados vinculado ao recurso por meio de algum tipo de vínculo, como uma URL; (II) modelo de metadados incorporado, no qual os metadados descritivos são considerados parte das informações que os compõem. Nesse aspecto, tem-se como exemplo de metadados que podem ser usados para descrever objetos o MODS. Além disso, os metadados podem ser incorporados em página *web* pelos criadores, usando meta *tags* em HTML, como por exemplo as usadas em cabeçalhos – tipo <head> e </head> (Ramesh; Vivekavardhan; Bharathi, 2015).

Quadro 3 - Metadados descritivos aplicáveis à comunicação científica

Nº	Metadados	Detalhe e Fonte
41	AACR2	<i>Anglo-American Cataloguing Rules, 2nd edition</i> . Compêndio de regras para recursos bibliográficos. Descontinuada. Mantida pela American Library Association (ALA) ⁴¹ .
42	APPM	<i>Archives, Personal Papers, Manuscripts</i> . Padrão catalográfico para materiais de arquivo. Mantido pela Library of Congress e Society of American Archivists ⁴² .
43	BIBFRAME	<i>Bibliographic Framework</i> . Esquema que substitui o formato MARC. Para intercâmbio bibliográfico entre sistemas e aplicativos. Mantido pela Library of Congress ⁴³ .
44	CCO	<i>Cataloging Cultural Objects</i> . Para catalogação de objetos culturais, incluindo artefatos, documentos e obras de arte. Mantido pelo Visual Resources Association (VRA) ⁴⁴ .
45	CDWA	<i>Categories for the Description of Works of Art</i> . Padrão descritivo de obras de arte. Mantido pelo Getty Research Institute's Art and Architecture Thesaurus (AAT) ⁴⁵ .
46	DACS	<i>Describing Archives: A Content Standard</i> . Para descrever materiais de arquivo, desenvolvido pela Society of American Archivists (SAA) ⁴⁶ .
47	Darwin Core	Padrão da <i>Darwin Core Maintenance Interest Group</i> para descrição sobre biodiversidade ⁴⁷ .
48	DCMI	<i>Dublin Core Metadata Initiative</i> . Organização responsável por padrões interoperáveis para recursos digitais ⁴⁸ .
49	DCRM	<i>Descriptive Cataloging of Rare Materials</i> . Para descrever documentos ou materiais raros ⁴⁹ .
50	EAD	<i>Encoded Archival Description</i> . Descrever documentos arquivísticos, baseado em XML. Mantido pela Library of Congress e Society of American Archivists ⁵⁰ .
51	FRBR	<i>Functional Requirements for Bibliographic Records</i> . Modelo bibliográficos. Desenvolvido pela International Federation of Library Associations and Institutions (IFLA) ⁵¹ .

41 Disponível em: <https://bit.ly/3VaYY5t>.

42 Disponível em: <https://www2.archivists.org/node/10163>.

43 Disponível em: <https://www.loc.gov/bibframe/>.

44 Disponível em: <https://bit.ly/3Ls85M0>.

45 Disponível em: <https://bit.ly/40QyqI9>.

46 Disponível em: <https://bit.ly/2UjlqMY>.

47 Disponível em: <https://dwc.tdwg.org/>.

48 Disponível em: <https://www.dublincore.org/>.

49 Disponível em: <https://bit.ly/41UfZ69>.

50 Disponível em: <https://bit.ly/41TOPMD>.

51 Disponível em <https://bitlybr.com/EOAzR>.

Nº	Metadados	Detalhe e Fonte
52	IFLA-LRM	<i>IFLA Library Reference Model</i> . Modelo bibliográfico. Agregou FRBR. Mantido pela IFLA ⁵² .
53	ISAD(G)	<i>General International Standard Archival Description</i> . Descrição de coleções arquivísticas. Influenciou o esquema EAD e SEPIADES europeu ⁵³ .
54	MARC	<i>Machine Readable Catalogue</i> . Formato de intercâmbio bibliográfico. Mantido pela Library of Congress ⁵⁴ .
55	MODS	<i>Metadata Object Description Schema</i> . Para descrever recursos digitais. Baseado em XML. Mantido pela Library of Congress ⁵⁵ .
56	RDA	<i>Resource Description and Access</i> . Norma catalográfica, substitui a AACR2. Mantido pela RDA Steering Committee ⁵⁶ .
57	SEPIADES	<i>SEPIA Data Element Set</i> . Para descrever coleções fotográficas ⁵⁷ .
58	VRA Core	Padrão baseado no CDWA. Descreve obras de arte e imagens culturais. Mantido pela Library of Congress e Visual Resources Association ⁵⁸ .

Fonte: elaborado pelos autores.

Padrões de metadados: princípios

Observa-se que os metadados administrativos abrangem aspectos técnicos e operacionais relacionados ao uso do recurso. Também se relacionam com metadados de preservação e de direitos, como, por exemplo, para informar se o recurso está disponível sob alguma licença Creative Commons. Os metadados estruturais organizam internamente o recurso, estabelecendo diretrizes de hierarquia e ligação entre seus componentes, como, por exemplo, a ordem de capítulos de um livro. Os metadados descritivos são responsáveis pela descrição e pelo contexto do recurso, facilitando sua identificação,

52 Disponível em: <https://www.ifla.org/resources/?oPubId=11412>.

53 Disponível em: <https://bit.ly/2nP1cZ9>.

54 Disponível em: <https://www.loc.gov/marc/>.

55 Disponível em: <https://www.loc.gov/standards/mods/>.

56 Disponível em: <https://bit.ly/3Narli3>.

57 Disponível em: <https://bit.ly/3L2ln0k>.

58 Disponível em: <https://www.loc.gov/standards/vracore/>.

como títulos, autorias, datas e locais de publicação. Os três tipos de metadados operam em conjunto e desempenham funções complementares.

É comum encontrar exemplos de metadados que são relevantes para uma variedade de publicações e dados de pesquisas científicas e desempenham um papel fundamental na composição de um acervo documental e/ou bibliográfico. Aderem a isso os Identificadores de Objeto Digital (DOIs), por exemplo, que permitem fácil localização de conteúdo digital, e que estão vinculados a outros metadados incorporados em uma produção científica. Isso pode incluir etiquetas de identificação como International Standard Serial Number (ISSN) e International Standard Book Number (ISBN). Esses metadados são usados por programas como o Altimetric (2023) para rastrear atividades *on-line* relacionadas a uma publicação exclusiva. No entanto, os processos atuais ainda não estão usufruindo de todo o potencial dos metadados no cenário em constante evolução das comunicações acadêmicas (Metadata, 2020).

Em um artigo sobre o tema, Kaiser e demais autores (2020) apresentam princípios fundamentais para o uso de metadados na comunicação científica, que são: (1) compatibilidade: fornece orientações tanto para máquinas como para pessoas; (2) metadados devem ser abertos, interoperáveis – permitir intercâmbio de informações –, analisáveis – conter caracteres reconhecíveis e analisáveis por computador –, acionáveis por máquina – estruturados para reconhecimento por *software* –, e legíveis por humanos; (3) Identificadores Persistentes (PIDs), como o ORCID ID, devem ser usados para garantir o suporte à interoperabilidade e a referência confiável às entidades digitais; ao contrário, um endereço de URL não é um PID, por estar suscetível a quebras de endereço e indisponibilidade a longo prazo; (4) ser completo e abrangente, se possível para refletir o conteúdo, os componentes e os relacionamentos publicados; (5) credível: os metadados devem ser de procedência clara, confiável e precisa; (6) curadoria: refletir atualizações e incorporar novos elementos. Esse conjunto de princípios também são conhecidos como *Findable, Accessible, Interoperable, Reusable* (FAIR) (Sales *et al.*, 2021).

É observado, pelos autores, que os metadados estão em constante evolução e, portanto, devem ser mantidos e atualizados ao longo do tempo. Esses princípios visam fornecer uma compreensão sólida em torno dos

metadados necessários para apoiar a comunicação científica. A linguagem dos metadados pode ser específica para uma determinada disciplina. Assim, as abordagens de melhores práticas devem permitir flexibilidade no uso de metadados entre as comunidades. Além disso, os padrões existentes devem ser considerados e reutilizados tanto quanto possível para favorecer a interoperabilidade entre esquemas de metadados, minimizando a redundância.

Não há um único padrão de metadados que atenda à diversidade de formatos ou objetos digitais nas diferentes comunidades de usuários. Os metadados são projetados para finalidades específicas, nas quais o significado de cada elemento é definido pelas regras de conteúdo. Os quadros apresentados (1, 2 e 3) fornecem alguns exemplos de metadados que exemplificam a diversidade dos padrões.

Estudiosos, em geral, preferem publicar seus trabalhos em revistas eletrônicas, repositórios de *preprints* etc. As comunidades e provedores de conteúdo nas áreas das ciências em geral, do patrimônio cultural – biblioteca, museus e arquivos –, comércio eletrônico e gestão de negócios estão em constante busca por refinamentos e acordos na semântica dos metadados especializados, para permitir novos formatos de uso internacional.

Bibliotecas, centros de informação, arquivos e museus digitalizam as informações de patrimônio cultural e científico, exigindo a criação de metadados capazes de organizar e gerenciar esses recursos. Dessa forma, acrescida pela definição das tipologias de metadados, considera-se que não existe um tipo único de metadados aplicável a todos os recursos de informação, ou que atenda a toda comunidade científica ou de informação.

Como indicado nos quadros, há dezenas de metadados que cumprem funções variadas na gestão, identificação, preservação e descrição dos conteúdos ou de sua apresentação, ou característica técnica de produção. A relação de metadados apresentada ilustra, além da diversidade de padrões desenvolvidos para o tratamento e para a disseminação das informações, as especificações e especialidade dos profissionais envolvidos no desenvolvimento e uso de metadados, que às vezes não se destinam exclusivamente aos padrões de recuperação da informação. Algumas ferramentas são disponibilizadas para que “não especialistas” insiram os seus dados em campos pré-definidos com diretrizes específicas para a organização. Mesmo *softwares*,

como os editores de texto, executam automaticamente a análise do documento, extraem e geram valores por meio da incorporação de etiquetas de metadados, quando o documento é criado e quando ocorrem modificações neste. Um dos principais desafios na gestão e aplicação de metadados é identificar os elementos necessários para descrever um recurso e que são denominados como elementos essenciais. Esses elementos identificados devem descrever a origem, a composição e navegação, a confidencialidade, a legitimidade e o reconhecimento, a qualidade e longevidade, além dos termos e condições de uso dos objetos digitais.

Metadados aplicados na comunicação científica

Como salientado, os metadados configuram-se como informações estruturadas que possibilitam que qualquer resultado de pesquisa seja detectável e recuperável. Eles informam aos mecanismos de pesquisa dados sobre o conteúdo que é buscado (University of Toronto Press, 2015).

Os metadados, nesse contexto, são as informações que acompanham as etapas e resultados da pesquisa. Comum à maioria das pesquisas científicas são os elementos de metadados, como: autor, data, título, assunto, idioma e identificador padrão (Gregg *et al.*, 2019). Por essa razão, a adoção de um padrão de metadados reúne os requisitos fundamentais que definem os significados e a semântica dos dados para uma correta armazenagem, recuperação, interpretação e uso (Riley, 2017).

Salienta-se, na perspectiva de relevância dos metadados abertos para a produção e comunicação científica, a iniciativa do Metadata 20/20. Uma ação de colaboração, lançada em 2017, para promover metadados mais ricos, conectáveis e reutilizáveis para todos os resultados de pesquisa. Essa ação envolve organizações e indivíduos relacionados com as comunicações acadêmicas – editores, agregadores, provedores de serviços, bibliotecários, financiadores e os próprios pesquisadores. Destaque como membros dessa iniciativa as organizações ORCID, Crossref, DataCite, OpenAIRE, California Digital Library, Wikimedia e OCLC (Metadata, 2020).

Como resultado foi publicado “Metadata 2020 Principles”, um conjunto de princípios com base nas melhores práticas existentes como forma

de apoiar e complementar os princípios de dados FAIR, citado anteriormente. É da expectativa dessas organizações para com a iniciativa Metadata 20/20 o comprometimento em melhorar a qualidade e a interoperabilidade de seus metadados e priorizar, no âmbito da comunidade de pesquisa, os metadados enriquecidos por favorecerem a descoberta e a inovação. Avaliar e melhorar os metadados criados e consumidos e o que pode ser alcançado quando a comunidade trabalha colaborativamente em direção a um objetivo comum. Permitir e valorizar o intercâmbio de dados entre sistemas e comunidades; bem como sua reutilização para eliminar a duplicação de esforços (Meadows, 2017).

Metadados: exemplo aplicado aos repositórios de teses e dissertações

Metadados são encontrados em uma variedade de formatos e contextos, abrangendo além do campo científico. Eles são aplicados em análises de mercado, agências de classificação, críticas, fóruns de discussão, *blogs*, entre outros. O acesso aos metadados de repositórios institucionais enfrenta limitações por diferentes razões e é fundamental desenvolver mecanismos eficazes para esses ambientes. Nesses serviços, os dados são dinâmicos, e podem ser acessados por meio de diferentes mecanismos externos de coleta da *web*. No âmbito acadêmico, a existência de repositórios de teses e dissertação em uma universidade reflete a qualidade intelectual da produção científica dos seus acadêmicos. Além de também indicar a capacidade da instituição de liderar seus estudantes e docentes pesquisadores, bem como apoiar os trabalhos científicos originais. Esse é o caso da Networked Digital Library of Theses and Dissertations (NDLTD).

Networked Digital Library of Theses and Dissertations (NDLTD)

A NDLTD é uma organização internacional dedicada à adoção, ao uso, à disseminação e à preservação eletrônica de teses e dissertações. Sua proposta ocorre em 1987 e se estabelece em 1991, como resultado da criação de uma base de dados de Dissertações e Teses Eletrônicas (ETDs) em rede. Por meio de um sistema de pesquisa por instituição, data, autor, título, palavras-chave e texto completo, é possível acessar o conteúdo das instituições

participantes. É uma inovação frente à criação das bibliotecas digitais, ou repositórios digitais. Esse modelo representa um avanço no apoio à publicação eletrônica e ao acesso aberto por facilitar o compartilhamento em escala global (Networked Digital Library of Theses and Dissertations, 2023).

Para Yiotis (2008), a principal missão do NDLTD é ajudar os pós-graduandos a se tornarem alfabetizados em informações, tornando o ETD um requisito institucional. Para o NDLTD, o propósito da atividade ETD é o de preparar as gerações de estudiosos para atuar com eficiência como trabalhadores do conhecimento na Era da Informação.

Moxley (2001) previu um futuro no qual os pesquisadores iriam desafiar a concepção de escrita acadêmica. O texto linear com margens de uma polegada daria lugar à escrita hipertextual, *streaming* multimídia, espaços de interação virtual, modelagem tridimensional, por exemplo. Aspectos presentes no estágio atual da comunicação científica.

O projeto NDLTD define um padrão de metadados usados para descrever uma tese ou dissertação eletrônica. As instituições que lidam com essas monografias desenvolveram seus próprios padrões ou adaptaram os existentes. Esses padrões descrevem o autor, o documento e o seu contexto de produção de maneira a ser útil ao pesquisador, ao bibliotecário ou à equipe técnica responsável pela manutenção do documento em seu formato eletrônico. As teses e dissertações configuram, portanto, em ETDs, que são basicamente um subconjunto do repositório digital local de uma instituição (Hickey; Pavani; Suleman, 2008).

O Repositório Institucional (RI) é definido como “[...] um arquivo digital do produto intelectual criado pelo corpo docente, equipe de pesquisa e alunos de uma instituição e acessível a usuários finais dentro e fora da instituição, com poucas ou nenhuma barreira de acesso” (Crow, 2002).

Ter uma biblioteca digital, que a maioria das grandes universidades de pesquisa possui, torna fácil instituir um repositório ETD. Isso inclui serviços digitais, coleções digitais e iniciativas de pesquisa e desenvolvimento. O sucesso do programa ETD ajudou a criar uma conscientização acadêmica nas universidades. Os métodos estabelecidos pelo NDLTD reduziram a dificuldade de desenvolver do zero o conhecimento tecnológico necessário para ensinar docentes, estudantes, gestores e bibliotecários sobre ETDs.

O NDLTD adota como padrão de ETD os elementos de metadados comuns do Dublin Core (DC) – acrescido de um novo elemento específico para teses (subelemento). Fornece diretrizes sobre quais informações relacionadas a uma ETD pertencem a cada elemento, elencado no Quadro 4. Se um elemento DC mais geral for descrito como obrigatório (O), ele deve ser especificado completamente, mesmo que elementos qualificados sejam especificados. Se um elemento mais geral for repetível (R), deve-se presumir que os subelementos são repetíveis (R). Se um elemento contém texto livre, ele deve ser repetível para permitir ETDs que forneçam metadados em mais de um idioma. Além disso, também pode haver elemento geral ou subelemento opcional (Op) (Networked Digital Library of Theses and Dissertations, 2023).

Quadro 4 - Padrão de metadados DC especificado pelo NDLTD

Elemento DC	Subelemento - qualificador	Descrição	Opção
dc.title		Nome do recurso. Título da obra ou equivalente	O R
	dc.title.alternative	Título alternativo da tese ou dissertação	Op R
dc.creator		Responsável pelo conteúdo do recurso. Nas teses ou dissertações, refere-se ao autor da obra. É inserido como aparece na obra ou equivalente, com um <i>link</i> para um registro de autoridade, se disponível.	O R
dc.subject		Assunto do conteúdo. Nas teses e dissertações, as palavras-chave ou assuntos listados. O qualificador do esquema é usado como vocabulário controlado.	O R
dc.description		Descrição do conteúdo do recurso. Nas teses e dissertações é o texto do resumo. O <i>dc.description</i> é intercambiável com o elemento <i>dc.description.abstract</i> .	O R
	dc.description.abstract	Texto completo do resumo.	Op R
	dc.description.note	Informações adicionais sobre a tese ou dissertação.	Op R
	dc.description.release	Descrição da versão da obra. Usado em erratas etc.	Op R
dc.publisher		Responsável por disponibilizar o recurso. Pode ou não ser o mesmo que <i>thesis.degree.grantor</i> . Deve ser inserido conforme aparece na obra ou recurso, com um <i>link</i> para um registro de autoridade, se disponível.	Op R

Elemento DC	Subelemento - qualificador	Descrição	Opção
	dc.publisher.country	O país em que o documento foi originalmente publicado.	
dc.contributor		Responsável pelas contribuições ao conteúdo. Usado para coautores, consultores e membros de comitês. Coautores da obra no todo usar <i>dc.creator</i> .	Op R
	dc.contributor.role	Alguém responsável pela criação ou aprovação da obra. Exemplo: conselheiro, membro do comitê, árbitro, jurado.	Op
dc.date		Data associada ao ciclo de vida do recurso. Registrar segundo a ISO 8601.	O
dc.type		Gênero do conteúdo. Distingue o recurso de outros e identifica os tipos de conteúdo incluídos. A frase “tese ou dissertação eletrônica” é recomendada como um valor para este elemento. Tipo de conteúdo é indicado por vocabulário padrão ⁵⁹ . Grau e nível de educação se controlam em <i>thesis.degree</i> .	O R
dc.format		Manifestação física ou digital do recurso. Na tese ou dissertação, informa o formato eletrônico em que a obra é armazenada. Indicar “desconhecido” se nenhuma informação de formato for disponível.	Op R
dc.identifier		Referência inequívoca ao recurso. Para fornecer a URI onde a obra pode ser visualizada. URNs persistentes como PURLs ⁶⁰ ou Handles ⁶¹ são recomendados.	O R
dc.language		Idioma do conteúdo do recurso. Deve ser o idioma principal no qual a obra é registrada. Partes da obra maior que aparecem em outros idiomas devem usar o qualificador lang. Para nomes de idiomas usar ISO 639-2 (ou RFC 1766).	Op R
dc.coverage		Extensão do conteúdo. Usado para períodos temporais ou regiões espaciais. Outro tipo de “coverage”, usar <i>dc.subject</i> .	Op R
dc.rights		Informações sobre direitos autorais. Condições nas quais a obra pode ser distribuída, reproduzida etc. Há três níveis válidos: 0) não acessível ao público; 1) acesso público limitado; 2) acessível ao público.	Op R
thesis.degree			
	thesis.degree.name	Nome do grau associado, como aparece na obra. Grau de escolaridade associado ao documento.	Op R

59 Disponível em: <http://dublincore.org/>.

60 Disponível em: <http://purl.org/>.

61 Disponível em: <http://handle.net/>.

Elemento DC	Subelemento - qualificador	Descrição	Opção
	thesis.degree.level	Nível de escolaridade: 0) graduação (pré-mestrado); 1) mestrado (pré-doutorado); 2) doutorado – inclui pós-doutorado.	Op R
	thesis.degree. discipline	Área de estudo do conteúdo intelectual do documento. Em geral, será o nome de um programa ou departamento.	Op R
	thesis.degree. grantor	Instituição que confere o grau. Como outros nomes e instituições.	Op R

Fonte: adaptado de Hickey, Pavani e Suleman (2008, tradução nossa).

Além dos atributos especificados no Quadro 4, para os elementos de metadados DC, também há alguns atributos globais que podem ser especificados para qualquer elemento. Exemplo: (1) *lang*: indicação do idioma no qual o valor de um campo é escrito. Isso é separado do elemento *dc.language*, que indica o idioma principal da obra; (2) *translated*: elemento que indica se o valor de um campo é uma tradução fornecida por alguém que não seja o autor. Para traduções fornecidas pelo autor, indica-se o idioma do próprio campo – usar o atributo “*lang*”; (3) *scheme*: descrição do vocabulário ou esquema usado para determinar o(s) assunto(s).

A padronização de metadados em repositórios de teses e dissertações favorece iniciativas como a La Referencia – Red de repositórios de acceso abierto a la ciência. A La Referencia é uma rede latino-americana de repositórios de acesso aberto que nasceu por meio de um acordo de cooperação assinado em Buenos Aires em 2012, que reflete a vontade política de disponibilizar em acesso aberto a produção científica da América Latina como um bem público regional, com ênfase nos resultados financiados com fundos públicos. Hoje já fazem parte da rede a Argentina, Brasil, Chile, Costa Rica, Colômbia, Equador, El Salvador, Espanha, México, Panamá, Peru e Uruguai, e somam mais de 4 milhões de documentos (La Referencia [...], [20--]). Essa forma de agregação da produção científica da região fortalece a colaboração e o compartilhamento entre os países participantes.

Considerações finais

Os metadados são parte integrante da comunicação científica, por serem recursos chave para localização, organização e recuperação da informação. A quantidade de esquemas e padrões existentes indica que não há um modelo único de metadados suficiente para descrever a diversidade documental ou que atenda à variedade de aplicações e usos. Muitas são as iniciativas que reúnem especialistas, editores, autores, bibliotecários e demais membros da comunidade para a construção de padrões específicos que possam resolver questões funcionais de interoperabilidade, organização, arquivamento e preservação em rede.

Embora o padrão de metadados DC apresente uma flexibilidade de aplicação universal, com integração com outros formatos e padrões atuais, ele ainda não é abrangente o suficiente para atender a registros mais detalhados e complexos. Entretanto, a multiplicidade de padrões de metadados resulta na necessária gestão das estruturas existentes. Além disso, recomenda-se, como princípio de padronização de metadados, que sejam flexíveis o bastante para permitir a exportação por meio da interoperabilidade dos dados, com base na compatibilidade, abrangência, confiabilidade e curadoria. Assim como os metadados são dinâmicos, necessitam de constante atualização, à medida em que surgem novos recursos informacionais e tecnológicos. Assim, é essencial o acompanhamento dessas mudanças e de adaptação das estruturas de metadados para garantir sua eficácia e relevância na comunicação científica.

Referências

ALTIMETRIC. About us. *Altmetric*, 2023. Disponível em: <https://www.altmetric.com/>. Acesso em: 14 abr. 2023.

BAGLEY, P. R. *Extension of programming language concepts*. Philadelphia: University City Science Center, 1969. Disponível em: <https://web.archive.org/web/20161114233132/http://www.dtic.mil/dtic/tr/fulltext/u2/680815.pdf>. Acesso em: 22 abr. 2023.

BAGLEY, S. *Philip R. Bagley bibliography*. [S. l.: s. n.], 2014. Disponível em: <https://stevenbagley.net/blog/files/prb-bibliography.pdf>. Acesso em: 22 abr. 2023.

BAPTISTA, F.; FERNEDA, E. O padrão brasileiro de metadados para teses e dissertações MTD-BR: aspectos e contribuições. In: SEMINÁRIO EM CIÊNCIA DA INFORMAÇÃO, 6., 2016, Londrina. *Anais [...]*. 2016. Londrina: UEL, 2016. Disponível em: <https://www.uel.br/eventos/cinf/index.php/secin2016/secin2016/paper/viewFile/323/202>. Acesso em: 10 mar. 2023

BARBALHO, C. R. S. Periódico científico: parâmetros para avaliação de qualidade. In: FERREIRA, S. M. S. P.; TARGINO, M. Das G. (org.). *Preparação de revistas científicas: teoria e prática*. Rio de Janeiro: Reichmannn & Autores Editores, 2005.

BASUMALLICK, C. What is Metadata? Definition, Types, Uses, and Example. *Spiceworks*, [s. l.], 20 Oct. 2022. Disponível em: <https://www.spiceworks.com/tech/devops/articles/what-is-metadata/>. Acesso em: 27 abr. 2023.

BULLETIN OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY. [Silver Spring: s. n.], v. 24, n. 1, p. 4-29, Oct./Nov.1997.

CRANE, D. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. Chicago: University of Chicago, 1972.

CROW, R. *The case for institutional repositories: a SPARC position paper*. Washington: the Scholarly Publishing: Association of Research Libraries, 2002. Disponível em: <https://web.archive.org/web/20050205085442/www.arl.org/sparc/IR/ir.html>. Acesso em: 10 maio 2023.

FREDRIKSSON, E. H. *A Century of Science Publishing*. Amsterdam: IOS Press, 2001.

GREGG, W. J. *et al.* A literature review of scholarly communications metadata. *Research Ideas and Outcomes*, [s. l.], v. 5, 5 Aug. 2019. DOI 10.3897/rio.5.e38698. Disponível em: <https://riojournal.com/article/38698>. Acesso em: 23 abr. 2023.

GUERRINI, M. Metadata: The Dimension of Cataloging in the Digital Age. *Cataloging & Classification Quarterly*, [Philadelphia], v. 60, n. 5,

p. 411-423, 2022. Disponível em: https://www.researchgate.net/publication/361882057_Metadata_The_Dimension_of_Cataloging_in_the_Digital_Age. Acesso em: 20 fev. 2023.

HICKEY, T.; PAVANI, A.; SULEMAN, H. (ed.). *ETD-MS v1.1: an Interoperability Metadata Standard for Electronic Theses and Dissertations*. [S. l.: s. n.], 2008. Disponível em: <https://ndltd.org/wp-content/uploads/2021/04/etd-ms-v1.1.html>. Acesso em: 10 maio 2023.

IOWA STATE UNIVERSITY. *LIB 160: Information Literacy: Iowa State University Library Instruction Services*. [S. l.]: Iowa State University, 2019. Disponível em: <https://iastate.pressbooks.pub/lib160info1ed/>. Acesso em: 22 abr. 2023.

KAISER, K. *et al.* Methods & Proposal for Metadata Guiding Principles for Scholarly Communications. *Research Ideas and Outcomes*, [s. l.], v. 6, maio 2020. DOI 10.3897/rio.6.e53916. Disponível em: <https://riojournal.com/article/53916/>. Acesso em: 10 abr. 2023.

LEARN MORE. *Metadata*, [s. l., 20-]. Disponível em: <https://metadata2020.org/learn-more/>. Acesso em: 14 abr. 2023.

LEIPZIG, J. *et al.* The role of metadata in reproducible computational research. *Patterns*, [s. l.], v. 2, p. 1-28, Sept. 2021.

LYMAN, P. Digital Documents and the Future of the Academic Community. In: EKMAN, R.; QUANDT, R. E. (ed.). *Technology and scholarly communication*. Los Angeles: University of California Press, 1997. *E-book*. Disponível em: <https://bit.ly/3KZDbJo>. Acesso em: 18 abr. 2023.

MEADOWS, A. Metadata 2020: What, Why, Who, How? *ORCID*, Sept. 2017. Disponível em: <https://info.orcid.org/metadata-2020-what-why-who-how/>. Acesso em: 3 maio 2023.

MÉNDEZ RODRÍGUEZ, E. M. *Metadados y recuperación de información: padrões, problemas e aplicabilidades em bibliotecas digitais*. 2001. Thesis (Doctorado en Documentación) – Departamento de Biblioteconomía y Documentación, Universidad Carlos III de Madrid, Getafe, 2001.

METADATA Management: What is the difference between Metadata and Data? *ZEENEA*, [s. l.], 2023. Disponível em: <https://zeenea.com/the-difference-between-metadata-data/#:~:text=The%20main%20>

difference%20between%20Data,more%20context%20for%20data%20users.
Acesso em: 20 abr. 2023.

METADATA in scholarly communications. *The Publication Plan*, [s. l.], July 2020. Disponível em: <https://thepublicationplan.com/2020/07/03/metadata-in-scholarly-communications/>. Acesso em: 14 abr. 2023.

MOXLEY, J. M. Universities should require electronic theses and dissertations. *Educause Quarterly*, [s. l.], v. 3, p. 61-63, 2001, Disponível em: www.educause.edu/ir/library/pdf/EQM0139.pdf. Acesso em: 10 abr. 2023.

MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. *Ciência da Informação*, Brasília, DF, v. 35, n. 2, p. 27-38, maio/ago. 2006. Disponível em: <https://www.scielo.br/j/ci/a/nGD3MkKfNxtjnnWshf3YVjP/?format=pdf&lang=pt>. Acesso em: 21 abr. 2023.

NATIONAL INFORMATION STANDARDS ORGANIZATION. *Understanding Metadata*. Bethesda: NISO, 2004. Disponível em: <https://repository.arizona.edu/bitstream/handle/10150/105486/niso.pdf?sequence=1&isAllowed=y>. Acesso em: 9 jan. 2022.

NETWORKED DIGITAL LIBRARY OF THESES AND DISSERTATIONS. *Mission, goals and history*. [S. l.]: NDLTD, 2023. Disponível em: <https://ndltd.org/mission-goals-and-history/>. Acesso em: 10 abr. 2023.

POMERANTZ, J. *Metadata*. Cambridge: MIT Press, 2015.

RAMESH, P.; VIVEKAVARDHAN, J.; BHARATHI, K. Metadata Diversity, Interoperability and Resource Discovery Issues and Challenges. *DESIDOC Journal of Library & Information Technology*, [s. l.], v. 35, n. 3, p. 193-199, May 2015. DOI 10.14429/djlit.35.3.8074. Disponível em: <https://publications.drdo.gov.in/ojs/index.php/djlit/article/view/8074/4884>. Acesso em: 20 abr. 2023.

LA REFERENCIA: Red Latinoamericana para la Ciencia Abierta. Argentina, [20--]. Disponível em: <https://www.lareferencia.info/es/>. Acesso em 20 ago. 2023.

RILEY, J. *Understanding Metadata: what is Metadata, and What is it For?* Bethesda: NISO, 2017. Disponível em: <https://www.fidgeo.de/fileadmin/>

user_upload/2016/07/2017_01-NISO-understanding-metadata.pdf.

Acesso em: 9 jan. 2022.

RUSSELL, J. M. La Comunicación científica a comienzos del siglo XXI. *Revista Internacional de Ciencias Sociales*, Paris, v. 53, n. 2, p. 271-282, 2001. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000123912_spa/PDF/123917spa.pdf.multi.nameddest=123912. Acesso em: 18 abr. 2023.

SALES, L. F. *et al.* (org.). *Princípios FAIR aplicados à gestão de dados de pesquisa*. Rio de Janeiro: IBICT, 2021. E-book. DOI 10.22477/9786589167242. Disponível em: https://ridi.ibict.br/bitstream/123456789/1182/2/IBICT_Principios%20FAIR%20aplicados%20a%20gest%c3%a3o%20de%20dados%20de%20pesquisa_2021.pdf. Acesso em: 19 maio 2023.

SAYÃO, L. F. Uma outra face dos metadados: informações para a gestão da preservação digital. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 15, n. 30, p.1-31, 2010. DOI 10.5007/1518-2924.2010v15n30p1. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2010v15n30p1/19527>. Acesso em: 9 jan. 2022.

SCHÖPFEL, J. Open access and document supply. *Interlending & Document Supply*, Boston, v. 42, n. 4, p.187-195, 2014. Disponível em: https://hal.science/sic_01083775. Acesso em: 21 abr. 2023.

SOUZA, V. B. P. de. *Dados massivos como apoio na identificação de informação institucional: estudo de caso na Universidade Federal de Rondônia* (UNIR). 2021. Tese (Doutorado em Ciência da Informação) - Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2021. Disponível em: <https://www.teses.usp.br/teses/disponiveis/27/27151/tde-02052022-115917/publico/ValmirBatistaPrestesdeSouzacorrigida.pdf>. Acesso em: 15 mar. 2023

UNIVERSITY OF TORONTO PRESS. *What is metadata?* Toronto, Jan. 2015. Disponível em: <https://blog.utpjournals.com/2015/01/27/what-is-metadata/>. Acesso em: 10 maio 2023.

YIOTIS, K. Electronic theses and dissertation (ETD) repositories: What are they? Where do they come from? How do they work? *OCLC Systems & Services: International Digital Library Perspectives*, [s. l.],

v. 24, n. 2, p. 101-115, 2008. Disponível em: https://www.researchgate.net/publication/220418606_Electronic_theses_and_dissertation_ETD_repositories_What_are_they_Where_do_they_come_from_How_do_they_work. Acesso em: 10 abr. 2023.

CAPÍTULO 10

Análise das ferramentas de geração automática e semiautomática de metadados: possibilidades e limitações

Jean Carlos Borges Brito

Dalton Lopes Martins

Introdução

A humanidade produz informações em volumes e variedades exponenciais diariamente, que são armazenadas principalmente em repositórios digitais. Com isso, conteúdos físicos legados estão sendo convertidos e mantidos em bancos de dados e unidades de armazenamento lógico, acompanhando a transformação digital atual por qual passam vários países do mundo, abrangendo governos, empresas e sociedade mundial.

Estudo publicado por Reinsel, Gantz e Rydning (2018) para o International Data Corporation (IDC) previa que o crescimento de dados aumentará de 45 *zettabytes* em 2019 para cerca de 175 *zettabytes* até 2025. Essa informação demonstra que em cinco anos, 6 bilhões de pessoas ou 75% da população mundial interagirão com dados todos os dias e cada pessoa conectada terá pelo menos uma interação com dados a cada 18 segundos. Conforme o IDC, grande parte da economia atual depende de dados e essa dependência só aumentará no futuro à medida que as entidades capturem, cataloguem, gerenciem e analisem os seus dados a partir de processos e tecnologias que aumentem a qualidade dos dados e permitam melhor exploração de seu valor agregado.

A utilização de abordagens de processos automatizados pode melhorar a eficiência das atividades de catalogação de dados nesse ambiente digital, disponibilizando em tempo real aquilo que é feito tradicionalmente de forma

manual. Segundo Mey e Silveira (2010), a catalogação pode ser entendida como a criação de conteúdos comunicativos que conectem mensagens de registros de conhecimento – reais ou *on-line* – com as necessidades dos usuários, por meio do estudo, da preparação e da organização das mensagens.

Cunha e Cavalcanti (2008, p. 70) ampliam essa interpretação discorrendo que “[...] a catalogação abrange não somente a descrição bibliográfica, mas também a análise temática com seus produtos, entre eles a identificação temática”. Um usuário potencial da informação é capaz de converter sua necessidade de informações em uma lista de referências para documentos armazenados e que contém informações úteis (Mooers, 1951). Nesse contexto, os metadados desempenham um papel essencial, pois descrevem os dados, facilitando sua compreensão e corroborando a eficácia de catalogação e sua obtenção.

Crystal e Land (2003) discorrem que para criar metadados para um milhão de documentos deveriam ser alocados 60 empregados por ano para realizar essa tarefa. É considerado um trabalho árduo, lento e caro se executado manualmente, continuam esses autores. Além disso, considerando que o conceito de documento e suas possibilidades de expressão midiática se expandem de forma significativa na era da *web*, torna-se proibitivo imaginar que a catalogação dos documentos seguirá continuamente sendo realizada apenas de forma manual.

Park e Brenza (2015) apresentam em seus estudos ferramentas de Geração Semiautomática de Metadados (GSAM), que dizem respeito ao uso de *software* para criação de registros de metadados com graus variados de supervisão por um especialista humano.

De acordo com Greenberg (2004), a Geração Automática de Metadados (GAM), baseada no conhecimento sobre indexação automática – associação de termos a documentos –, é mais eficiente, possui menor custo e é mais consistente do que processos executados por seres humanos. A autora enfatiza que segundo alguns pesquisadores, o meio mais efetivo para criar metadados é integrar métodos automáticos e semiautomáticos, um complementando o outro.

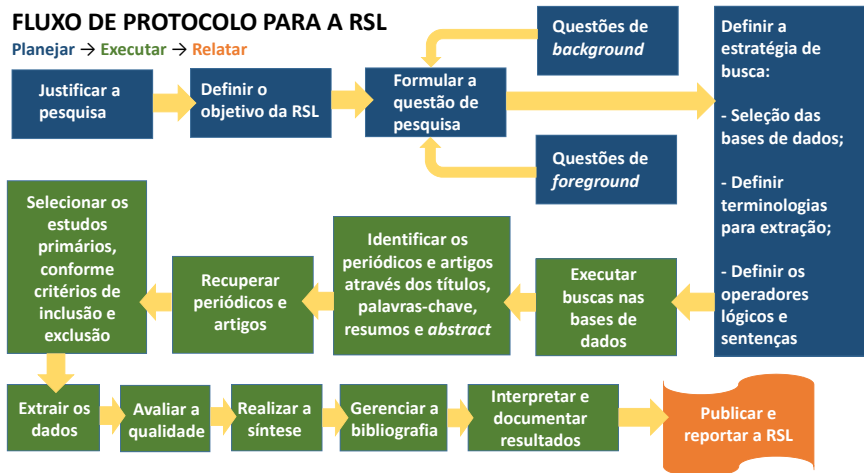
Diante dos fatos apresentados, este capítulo relata a realização de uma revisão sistemática da literatura sobre o tema GSAM, suas ferramentas,

técnicas, características, funções e aplicações. Executou-se uma pesquisa exploratória em bases de dados científicas, selecionando periódicos específicos para a avaliação de acordo com as principais listas de classificação no contexto da ciência da informação, identificando eventuais lacunas, a fim de sugerir novas áreas de investigação para auxiliar nas ações de catalogação e classificação dos dados.

A literatura sobre GSAM

A Revisão Sistemática de Literatura (RSL) se preocupa com a reprodutibilidade do método científico por outros pesquisadores, ou seja, promove a descrição explícita de cada passo do processo, demonstrando todas as atividades realizadas (Galvão; Ricarte, 2019). Neste estudo foi desenvolvido um protocolo de revisão, subdividido em três fases, conforme proposta descrita por Kitchenham (2004): planejar, executar e relatar.

Figura 1 – Protocolo da RSL



Fonte: adaptada de Kitchenham (2004).

Planejamento da RSL

O ponto de partida para a RSL foi a delimitação da questão de pesquisa, a partir da qual se definiu a população ou o problema, a intervenção, a comparação e o resultado (Galvão; Ricarte, 2019). Inicialmente, elencaram-se questões de *background* para fornecer a compreensão básica e conceitual do tema. Em seguida, para melhor definição do escopo, estabeleceram-se questões de *foreground*.

Quadro 1 – Questões de *background* e *foreground*

QUESTÕES DE <i>BACKGROUND</i>	QUESTÕES DE <i>FOREGROUND</i>
O que são metadados? Para que servem? O que é GSAM?	Quais técnicas, características, funções, ferramentas e aplicações da GSAM?

Fonte: dados da pesquisa.

Com o entendimento dos conceitos e da abrangência, formulou-se a seguinte questão de pesquisa para a investigação: “quais as possibilidades e limitações atuais das ferramentas de GSAM para o gestor de repositórios digitais?”.

Após a delimitação da questão de pesquisa, definiu-se as bases de dados que serão consultadas para obtenção de artigos. São apresentadas a seguir: Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (Brapci); Library and Information Science Abstract (Lisa); Library, Information Science & Technology Abstracts (Lista); Emerald Publishing Limited; Information Science and Technology Abstracts (Ista); Wiley Online Library; Web of Science e Scopus. Nesse ínterim, elaborou-se as seguintes terminologias para compor a extração dos dados, utilizando-se termos em português e inglês: “Geração automática de metadados”, “*Automatic metadata generation*”, “Geração semiautomática de metadados”, “*Semiautomatic metadata generation*”, “Ferramentas”, “*Tools*”, “Técnicas”, “*Techniques*”, “Características”, “*Features*”, “Funções”, “*Functions*”, “Aplicações” e “*Applications*”. Sabe-se que outros descritores poderiam trazer trabalhos eventualmente relacionados com o tema, como “indexação automática”, por

exemplo. No entanto, optou-se nesta pesquisa por analisar a literatura científica com foco no problema mais amplo, ou seja, a geração de metadados, sejam eles descritivos ou temáticos. Nos próximos trabalhos, pretende-se aprofundar em termos derivados.

Próximo passo, com o auxílio de um bibliotecário, foi a construção das sentenças de buscas utilizando os operadores lógicos, sendo definidas:

- Sentença em português: (((“Geração automática de metadados”) OR (“Geração semiautomática de metadados”)) AND (Ferramentas OR Técnicas OR Características OR Funções OR Aplicações));
- Sentença em Inglês: (((“Automatic metadata generation”) OR (“Semiautomatic metadata generation”)) AND (Tools OR Techniques OR Features OR Functions OR Applications)).

A partir dessas definições, realizaram-se atividades de pré-testes nas bases científicas para verificar se as sentenças deveriam passar por um processo de readequação, o que foi confirmado. Concluído o pré-teste, finalizou-se a fase de planejamento e após a aprovação por especialista desta etapa, passou-se à execução da pesquisa.

Execução da RSL

Para a execução da RSL, buscas foram executadas nas bases científicas definidas na etapa de planejamento. Ao acessar a página de cada periódico, realizou-se o preenchimento dos filtros de busca de forma a delimitar a recuperação da informação. Com isso, os seguintes critérios de inclusão e exclusão foram aplicados para obter as publicações:

Quadro 2 – Critérios de inclusão e exclusão

CRITÉRIOS DE INCLUSÃO	CRITÉRIOS DE EXCLUSÃO
Trabalhos científicos publicados entre os anos de 2010 e 2020	Trabalhos científicos publicados antes do ano de 2010
Descrição de estudo de caso, experimentos ou <i>survey</i> , outras RSLs	Título do trabalho não condizente com a proposta do projeto

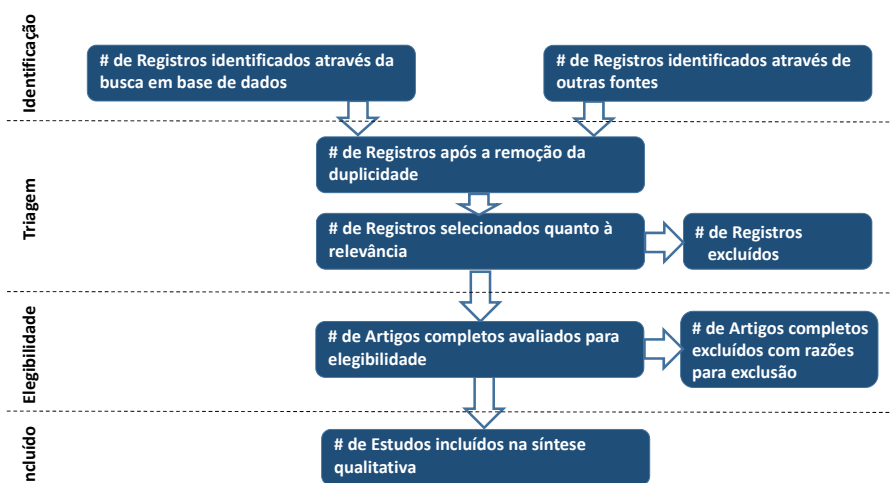
CRITÉRIOS DE INCLUSÃO	CRITÉRIOS DE EXCLUSÃO
Resumo ou <i>abstract</i> condizente com a proposta de pesquisa	Resumo ou <i>abstract</i> com fuga ao tema proposto na pesquisa
Artigos e periódicos	Publicações sem cunho científico
Publicações em inglês e português com disponibilidade completa e suporte em meio eletrônico	Disponibilização de partes da pesquisa, textos incompletos

Fonte: dados da pesquisa.

Duas bases de dados não retornaram resultado com as sentenças definidas na etapa de planejamento, sendo elas a Brapci e a Wiley Online Library. Assim, resolveu-se desmembrar as sentenças compostas em termos simples de busca, mas obteve-se zero resultado na recuperação de informação nesses dois repositórios. Interessante comentar que a Brapci, por ser uma base referencial em ciência da informação com artigos indexados desde 1972, não possui investigações relacionadas com o tema GSAM. Diante disso, constatou-se que os filtros do formulário dessa plataforma apresentavam falhas, pois ao utilizar qualquer termo de busca como teste, o critério de inclusão “Trabalhos científicos publicados entre os anos de 2010 e 2020” retornava consultas desde 1972 e não no período selecionado. Registra-se que essa ação foi realizada no período de 6 a 16 de maio de 2020.

Para analisar os critérios mínimos de qualidade será utilizado um conjunto de itens com base em evidências, baseado no modelo elaborado por Moher e demais autores (2009) denominado Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Prisma), obedecendo ao seguinte fluxograma:

Figura 2 – Fluxo de seleção de artigos, baseado no *framework* Prisma



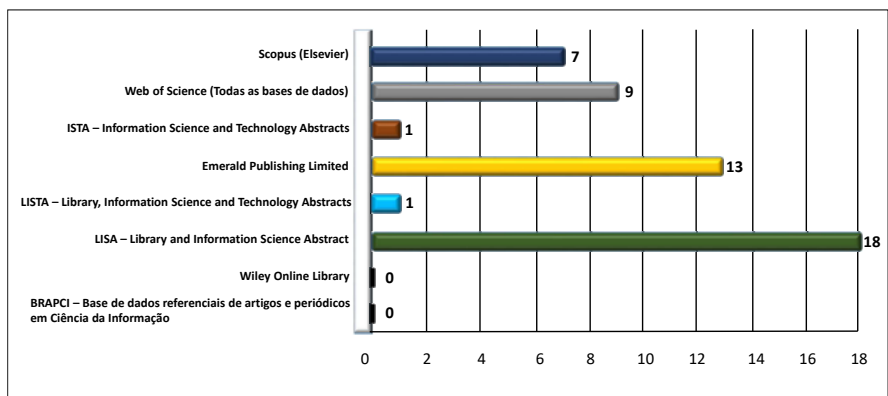
Fonte: adaptada de Moher e demais autores (2009).

As outras bases de dados pesquisadas retornaram o total de 49 trabalhos científicos, utilizando as sentenças definidas no planejamento.

A primeira atividade desta etapa consistiu na leitura do título e descrição dos registros identificados mediante a busca nas bases de dados. Foram identificados 15 artigos duplicados, ou seja, na busca executada retornaram resultados em mais de uma base de dados para o mesmo artigo.

A segunda atividade foi a leitura de resumo e palavras-chave, sendo excluídos 16 artigos por não serem úteis para continuidade da pesquisa. O total de 18 estudos foi selecionado para realização de análise. Entretanto, seis artigos completos foram excluídos após uma leitura preliminar, pois o conteúdo dos trabalhos não correspondia ao objeto deste projeto. Por fim, 12 estudos foram incluídos na síntese qualitativa.

Figura 3 – Quantidade de artigos nas bases pesquisadas (N = 49)



Fonte: dados da pesquisa (2021).

Executou-se o *download* do arquivo e o armazenado em repositório para fichamento e síntese.

Quadro 3 – Artigos selecionados para realização de fichamento e síntese

ANO	AUTORES	TÍTULO	NOTAS ADICIONAIS	INSTITUIÇÃO /LOCAL/PAÍS	CATEGORIA
2011	Kovacevic <i>et al.</i>	“Automatic extraction of metadata from scientific publications for CRIS systems”	Program: Electronic Library and Information Systems. v. 45 n. 4, p. 376-396	Novi Sad University / Novi Sad / Sérvia	Extração automática de metadados
2012	Maratea, A.; Petrosino, A.; Manzo, M.	“Automatic Generation of SCORM Compliant Metadata for Portable Document Format Files”	International Conference on Computer Systems and Technologies – CompSysTech	Parthenope University / Nápoles / Itália	Ferramenta de geração automática de metadados
2012	Verborgh <i>et al.</i>	“Enabling context-aware multimedia annotation by a novel generic semantic problem-solving platform”	Multimed Tools Appl 61, p. 105-129	Ghent University / Ghent / Bélgica	Ferramenta de geração automática de metadados

ANO	AUTORES	TÍTULO	NOTAS ADICIONAIS	INSTITUIÇÃO /LOCAL/PAÍS	CATEGORIA
2012	Sah, M; Wade, V.	“Automatic metadata mining from multilingual enterprise content”	Web semantics: Science, services and agents on the world wide web. v. 11, p. 41-62	Trinity College Dublin / Dublin / Irlanda	Extração automática de metadados
2013	Costa <i>et al.</i>	“EURAC SDI: A Near Real Time and Offline Automatic Metadata Generation Processing Chain”	GI Forum, Conference Proceedings. v. 1	Eurac Research / Bozen / Itália	Ferramenta de geração automática de metadados
2013	Vlachidis <i>et al.</i>	“Automatic Metadata Generation in an Archaeological Digital Library: Semantic Annotation of Grey Literature”	<i>In</i> : Przepiórkowski A. <i>et al.</i> (ed.). Computational Linguistics. Studies in Computational Intelligence, v. 458. Springer, Berlin, Heidelberg	University College London / Londres / United Kingdom	Ferramenta de geração automática de metadados
2015	Park, J.; Brenza, A.	“Evaluation of Semi-Automatic Metadata Generation Tools: A Survey of the Current State of the Art”	Information Technology and Libraries. v. 34, Ed. 3, p. 22-42	Drexel University / Filadélfia / EUA	Ferramenta de geração semiautomática de metadados
2015	Rafferty, J.; Nugent, C.; Liu, J.	“Automatic Metadata Generation Through Analysis of Narration Within Instructional Videos”	Transaction Processing Systems, J MedSyst n. 39, 94	Ulster University / Belfast / Irlanda do Norte	Ferramenta de geração automática de metadados
2018	Rodrigo <i>et al.</i>	“ScienceSearch: Enabling Search through Automatic Metadata Generation”	Conferência: 14th IEEE International Conference on EScience (E-Science). p. 93-104	Berkeley University / Califórnia / EUA	Ferramenta de geração automática de metadados

ANO	AUTORES	TÍTULO	NOTAS ADICIONAIS	INSTITUIÇÃO /LOCAL/PAÍS	CATEGORIA
2018	Yang, G.; Park, J.	“Automatic Extraction of Metadata Information for Library Collections”	International Journal of Advanced Culture Technology. v.6, n. 2, p. 117-122	Drexel University e Mokpo University / Filadélfia e Mokpo / EUA e Coréia do Sul	Ferramenta de geração automática de metadados
2019	Audichya, M. K.; Saini, J. R.	“Computational linguistic prosody rule-based unified technique for automatic metadata generation for Hindi poetry”	1st International Conference on Advances in Information Technology	Gujarat Technological University / Gujarat / Índia	Ferramenta de geração automática de metadados
2020	Morris, V.	“Automated Language Identification of Bibliographic Resources”	Cataloging & Classification Quarterly. v. 58, n. 1, p. 1-27	British Library / Wetherby / United Kingdom	Catálogo com ferramentas automatizadas

Fonte: dados da pesquisa.

Observando os artigos relacionados, nota-se que o assunto pesquisado é relevante e de interesse ao redor do mundo, pois se verificam investigações sobre o tema por diversos pesquisadores de várias universidades localizadas nos Estados Unidos, Europa e Ásia. A lacuna que se percebeu nesse levantamento foram pesquisas realizadas no contexto da ciência da informação no Brasil e na América Latina.

Realização da síntese

Foram criadas as categorias: extração automática de metadados, ferramenta de GAM, ferramenta de GSAM e catalogação com ferramentas automatizadas.

Kovacevic e demais autores (2011) apresentam um método para a extração automática de metadados de artigos científicos em formato PDF, que é projetado como parte integrante do sistema de informação para monitorar as atividades de pesquisa. O método é implementado como um complemento à entrada manual de metadados, no sentido de que os resultados da extração são oferecidos ao curador para inspecionar e corrigir antes de

armazená-los no repositório. O sistema é baseado nos métodos de aprendizado de máquina, ou seja, classificação e obteve-se melhor resultado com o uso do modelo Support Vector Machines.

Maratea, Petrosino e Manzo (2012) utilizaram algoritmos com técnicas de processamento de linguagem natural para GAM para conteúdo de aprendizagem. Aplicou-se como padrão uma coleção de especificações para o *e-learning* baseado na *web* amplamente adotado, denominado Modelo de Referência de Objeto Compartilhável para Conteúdo (Scorm). O objetivo desse modelo é permitir a interoperabilidade, o fácil acesso e a reutilização de unidades de aprendizagem baseadas na *web* para indústria, governo e universidade.

Verborgh e demais autores (2012) apresentam uma plataforma genérica de solução de problemas semânticos, que combina automaticamente os serviços da *web* para realizar uma tarefa predefinida e usa a *web* semântica como fonte de conhecimento para iniciar e manter ativamente o contexto da tarefa. Os autores realizaram a aplicação por meio de um caso de uso de anotação de imagem. Obtiveram resultados satisfatórios quanto à eficiência e à escalabilidade, sendo utilizada a ferramenta de raciocínio Eye para resolução de problemas semânticos e capaz de criar composições holísticas.

Sah e Wade (2012) investigam a utilização de ferramentas de GAM para fornecer informações avançadas de conteúdos acessados pelos usuários, fomentando aspectos de personalização do cliente, fazendo com que eles permaneçam mais tempo no *site*, incentivando-os a retornar ao provedor de serviços. Os autores desenvolveram uma ontologia DocBook e uma de recurso para extrair metadados estruturais e descritivos dos documentos DocBook no formato RDF.

Costa e demais autores (2013) apresentam em seu estudo a abordagem para GAM por intermédio de um método baseado em regras, codificado manualmente, implementado como *plug-ins* que geram metadados em formato padronizado extraído de um conjunto heterogêneo de dados geoespaciais. Os autores discorrem que a estação receptora do Museu Arqueológico de Bolzano (Eurac) recebe diariamente dados brutos das missões da Administração Nacional de Aeronáutica e Espaço (Nasa): Aqua, Terra e Suomi NPP. A pesquisa do instituto lida com muitos dados de satélite

diferentes: LANDSAT, RapidEye, ENVISAT e Quickbird. Eles executaram a ingestão e o manuseio de dados, fazendo uso de um aplicativo geral multi-tarefa centralizado, denominado Data Exchange Server (DES).

Vlachidis e demais autores (2013) realizam investigações sobre bibliotecas digitais, em especial a Europeana, tendo como objetivo de pesquisa executar a GAM com enriquecimento semântico significativo para seus objetos digitais vinculados, com o auxílio do Europeana Data Model (EDM), que resume o Cidoc Conceptual Reference Model (CRM), entre outros modelos de metadados. O enriquecimento semântico se caracteriza pelo processo de fornecimento de maior significado aos dados e metadados, auxiliando a integração, a compreensão e o processamento. Foi empregando o *kit* de ferramentas de Arquitetura Geral para Engenharia de Texto (Gate).

O artigo de Park e Brenza (2015) merece um destaque especial, pois essa publicação foi a mais indexada nas bases pesquisadas. Eles examinam uma variedade de ferramentas de geração semiautomáticas de metadados (N=39), analisando suas técnicas, recursos e funções. Esses autores desenvolveram uma matriz caracterizando cada ferramenta de GSAM analisada: nome, local *on-line*, técnicas usadas para geração de metadados (Greenberg, 2004; Polfreman; Broughton; Wilson, 2008), além de breve descrição das funções e recursos da ferramenta. O levantamento dos autores pode ser visualizado no Quadro 4 a seguir:

Quadro 4 – Ferramentas de geração semiautomática de metadados

FERRAMENTA	TÉCNICAS EMPREGADAS	CARACTERÍSTICAS E FUNÇÕES
ANVL/ERC Kernel Metadata Conversion Toolkit ¹	Colheita de <i>metatags</i> .	Um utilitário que pode automaticamente converter registros de formatos ANVL – A Name Value Language – para outros formatos, tais como XML, JSON – JavaScript Object Notation –, Turtle e Plain, entre outras.
Apache POI – Text Extractor ²	Extração de conteúdo; colheita de <i>metatags</i> ; geração automática extrínseca.	O Apache POI fornece extração básica de texto para todos os formatos de arquivo suportados pelo projeto. Além do texto – sem formatação –, ele pode acessar os metadados associados a um determinado arquivo, como título e autor.

1 Disponível em: <https://metacpan.org/pod/File::ANVL>.

2 Disponível em: <http://poi.apache.org/download.html>.

FERRAMENTA	TÉCNICAS EMPREGADAS	CARACTERÍSTICAS E FUNÇÕES
Apache Stanbol (movido para Apache Attic, desativado) ³	Extração de conteúdo; indexação automática.	Extrai metadados semânticos de arquivos PDF e de texto. Pode aplicar termos extraídos a ontologias.
Apache Stanbol (movido para Apache Attic, desativado) ⁴	Extração de conteúdo; indexação automática.	Extrai metadados semânticos de arquivos PDF e de texto. Pode aplicar termos extraídos a ontologias.
Apache Tika ⁵	Extração de conteúdo; colheita de <i>metatags</i> : geração automática extrínseca.	Construído no Apache POI, o Apache Tika toolkit detecta e extrai metadados e conteúdo de texto de vários documentos.
Ariadne Harvester ⁶	Colheita de <i>metatags</i> .	Uma colheitadeira de registros compatíveis com OAI-PMH que pode ser convertida em vários outros esquemas, como Metadados de Objetos de Aprendizagem (LOM).
Bibframe Tools ⁷	Colheita de <i>metatags</i> .	O BIBFRAME oferece várias ferramentas para a conversão de documentos MARCXML em documentos BIBFRAME. Serviço da <i>web</i> e <i>software</i> para <i>download</i> estão disponíveis.
Biblio Citation Parser ⁸	Extração de conteúdo.	Um conjunto de módulos para análise de citações.
CatMDEdit ⁹	Extração de conteúdo.	O CatMDEdit permite a criação automática de metadados para coleções de recursos relacionados, em particular séries espaciais que surgem como resultado da fragmentação de recursos geométricos em conjuntos de dados de tamanho gerenciável e escala semelhante.
CrossRef ¹⁰	Extração de conteúdo.	Este serviço da <i>web</i> retorna Identificadores de objetos digitais para referências inseridas.

3 Disponível em: <https://attic.apache.org/projects/stanbol.html>.

4 Disponível em: <https://attic.apache.org/projects/stanbol.html>.

5 Disponível em: <http://tika.apache.org/>.

6 Disponível em: <https://sourceforge.net/projects/ariadneksps/>.

7 Disponível em: <https://www.loc.gov/bibframe/implementation/>.

8 Disponível em: <https://metacpan.org/release/MJEWELL/Biblio-Citation-Parser-1.10>.

9 Disponível em: <https://inspire-reference.jrc.ec.europa.eu/tools/catmdedit><https://sourceforge.net/projects/catmdedit/>.

10 Disponível em: <https://doi.crossref.org/simpleTextQuery>.

FERRAMENTA	TÉCNICAS EMPREGADAS	CARACTERÍSTICAS E FUNÇÕES
Data Fountains ¹¹	Extração de conteúdo; indexador automático; colheita de <i>metatag</i> ; geração automática extrínseca.	Digitaliza documentos HTML e extrai primeiro as informações contidas nas <i>metatags</i> . Se as informações não estiverem disponíveis nas <i>metatags</i> , o programa usará outras técnicas para atribuir valores. Inclui um rastreador da <i>web</i> focado que pode segmentar <i>sites</i> sobre um assunto específico.
Digital Record Object Identification (DROID) ¹²	Geração automática extrínseca.	DROID é uma ferramenta de <i>software</i> desenvolvido pelo Arquivo Nacional (Reino Unido) para execução automática em lote de identificação de formatos de arquivo.
Dublin Core Meta Toolkit ¹³	Colheita de <i>metatags</i> .	Transforma dados coletados por diferentes métodos em metadados compatíveis com Dublin Core (DC).
Dspace ¹⁴	Colheita de <i>metatags</i> : Geração automática extrínseca; Marcação Social.	Extrai automaticamente informações técnicas sobre o formato e o tamanho do arquivo. Também pode extrair algumas informações das <i>metatags</i> .
Editor-Converter Dublin Core Metadata ¹⁵	Colheita de <i>metatags</i> : geração automática extrínseca.	Digitaliza documentos HTML, coletando metadados de <i>tags</i> e convertendo-os em DC.
Embedded Metadata Extraction Tool (EMET) ¹⁶	Extração de conteúdo; colheita de <i>metatag</i> ; geração automática extrínseca.	O EMET é uma ferramenta projetada para extrair metadados incorporados em arquivos JPEG e TIFF; detecta corrupção de arquivo de imagem e exporta resultados em arquivo Excel.

11 Disponível em: <https://sourceforge.net/projects/datafountains/>.

12 Disponível em: <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>.

13 Disponível em: <https://sourceforge.net/projects/dcmetatoolkit/>.

14 Disponível em: <https://duraspace.org/dspace/>.

15 Disponível em: <http://library.kr.ua/dc/dcreditunie.html>.

16 Disponível em: <https://sourceforge.net/projects/emet/>.

FERRAMENTA	TÉCNICAS EMPREGADAS	CARACTERÍSTICAS E FUNÇÕES
Firefox Dublin Core Viewer Extension (Plug-in descontinuado) ¹⁷	Colheita de <i>metatag</i> ; geração automática extrínseca.	Digitaliza documentos HTML, coletando metadados de <i>tags</i> e exibindo-os no DC.
FreeCite (descontinuado e substituído por AnyStyle) ¹⁸	Extração de conteúdo.	Ferramenta de análise gratuita para a extração de informações de referência. Pode ser baixado ou usado como um serviço <i>web</i> .
General Architecture for Text Engineering (Gate) ¹⁹	Extração de conteúdo; indexação automática.	Processador de linguagem natural e extrator de informações.
JHove ²⁰	Geração automática extrínseca.	Extrai metadados sobre o formato e tamanho do arquivo, bem como validando a estrutura do formato do arquivo identificado.
Kea ²¹	Extração de conteúdo; indexação automática.	Analisa os textos completos dos recursos e extrai frases-chave. As frases-chave também podem ser mapeadas para ontologias personalizadas ou vocabulários controlados para atribuição de termos do assunto.
MarcEdit ²²	Colheita de <i>metatag</i> .	Coleta dados compatíveis com OAI-PMH e os converte em vários formatos, incluindo DC e MARC.
MetaGen ²³	Extração de conteúdo; indexação automática.	O modelo de texto T4 para Visual Studio é uma mistura de blocos de texto e lógica de controle – fragmentos de código em Visual C# ou no Visual Basic – que pode gerar um arquivo de texto de qualquer tipo, como uma página da <i>web</i> , um arquivo de recurso ou um código-fonte do programa em qualquer idioma. O MetaGen usa o T4 para gerar metadados para projetos Silverlight.

17 Disponível em: <https://www.splintered.co.uk/experiments/73/>.

18 Disponível em: <https://anystyle.io/>.

19 Disponível em: <https://gate.ac.uk/overview.html>.

20 Disponível em: <https://jhove.openpreservation.org/>.

21 Disponível em: <http://community.nzdl.org/kea/>.

22 Disponível em: <https://marcedit.reeset.net/>.

23 Disponível em: <https://www.codeproject.com/Articles/41910/MetaGen-A-Project-metadata-Generator-for-Visual-St>.

FERRAMENTA	TÉCNICAS EMPREGADAS	CARACTERÍSTICAS E FUNÇÕES
MetaGenerator ²⁴	Extração de conteúdo.	Um <i>plug-in</i> que gera automaticamente descrição e <i>metatags</i> de palavras-chave puxando texto do conteúdo do Joomla. Com este <i>plug-in</i> é possível controlar algumas opções de título e adição de URL <i>metatags</i> .
Metatag Extractor Software ²⁵	Colheita de <i>metatag</i> .	Permite recursos personalizáveis de extração, coleta de <i>metatags</i> e informações de contato de sites.
My Meta Maker ²⁶	Colheita de <i>metatag</i> .	Permite converter dados inseridos manualmente em DC.
National Library of New Zealand – Metadata Extraction Tool ²⁷	Geração automática extrínseca.	Desenvolvido pela Biblioteca Nacional da Nova Zelândia para extrair programaticamente metadados de preservação de vários formatos de arquivo, como documentos PDF, arquivos de imagem, arquivos de som, documentos do Microsoft Office e outros.
Omeka ²⁸	Geração automática extrínseca; marcação social.	Extrai automaticamente informações técnicas sobre o formato e tamanho do arquivo.
Ont-O-Mat (descontinuado)	Extração de conteúdo.	Auxilia o usuário na anotação de sites compatíveis com a web semântica. Pode incluir um recurso que sugere automaticamente partes do site a serem anotadas.
Open Text Summarizer ²⁹	Extração de conteúdo.	Extrai frases pertinentes de um recurso para criar uma descrição de texto livre.
ParsCit ³⁰	Extração de conteúdo.	Pacote de análise de cadeia de código aberto para a extração de informações de referência de artigos acadêmicos.
Photo RDF-Gen ³¹	Colheita de <i>metatag</i> .	Gera saída DC e Resource Description Framework (RDF) a partir da entrada inserida manualmente.

24 Disponível em: <https://extensions.joomla.org/extension/site-management/seo-a-metadata/seo-generator/>.

25 Disponível em: <https://meta-tag-extractor.soft112.com/>.

26 Disponível em: <http://old.isn-oldenburg.de/services/mmm/>.

27 Disponível em: <http://meta-extractor.sourceforge.net/>.

28 Disponível em: <http://omeka.org/>.

29 Disponível em: <https://open-text-summarizer.soft112.com/>.

30 Disponível em: <https://parscit.comp.nus.edu.sg/>.

31 Disponível em: http://www.webposible.com/utilidades/photo_rdf_generator_en.html.

FERRAMENTA	TÉCNICAS EMPREGADAS	CARACTERÍSTICAS E FUNÇÕES
PyMarc ³²	Colheita de <i>metatag</i> .	Biblioteca <i>python</i> para trabalhar com dados bibliográficos codificados em MARC21. Ele fornece uma Application Program Interface (API) para leitura, gravação e modificação de registros MARC.
RepoMMan (Não disponível em 23 de julho de 2020) ³³	Extração de conteúdo; colheita de <i>metatag</i> ; geração automática extrínseca.	Extrai automaticamente vários elementos para documentos enviados ao Fedora, como autor, título, descrição e palavras-chave, entre outros. Os resultados são apresentados ao usuário para revisão.
Sherpa/RoMEO (descontinuada em julho de 2020) ³⁴	Colheita de <i>metatag</i> .	Uma API de máquina para máquina que permite a pesquisa e importação automáticas de editores e revistas.
Simple Automatic Metadata Generation Interface (Samgl) (descontinuada em 23 de julho de 2020)	Extração de conteúdo; geração automática extrínseca.	Um conjunto de ferramentas capaz de extrair automaticamente elementos de metadados como frase-chave e idioma dos documentos a partir do contexto em que um documento existe.
URL and Metatag Extractor ³⁵	Colheita de <i>metatag</i> .	Permite a pesquisa direcionada de <i>sites</i> e extrai URLs e <i>metatags</i> desses <i>sites</i> .
Termine ³⁶	Extração de conteúdo.	Extrai palavras-chave de textos através da análise do C-Value e Acrônimos, um identificador de sigla e dicionário. Disponível gratuitamente serviço <i>web</i> para uso acadêmico.
Yahoo Content Analysis API (descontinuado em 30 de junho de 2020)	Extração de conteúdo; Indexação automática.	Detecta entidades/conceitos, categorias e relacionamentos dentro de conteúdo não estruturado. Classifica as entidades/conceitos por sua relevância.

Fonte: Park e Brenza (2015) com atualizações e adaptações do autor.

Realizou-se acesso às ferramentas descritas no Quadro 4 no período de 11 de julho de 2020 a 26 de julho de 2020, sendo identificados alguns *links* que estavam quebrados – sendo todos atualizados. Algumas ferramentas foram ou serão descontinuadas em breve, outras possuem *link* para

32 Disponível em: <https://pypi.org/project/pymarc/>.

33 Disponível em: <http://www.ukoln.ac.uk/repositories/digirep/index/RepoMMan>.

34 Disponível em: <https://v2.sherpa.ac.uk/api/>.

35 Disponível em: <https://metatagsextractor.com/>.

36 Disponível em: <http://www.nactem.ac.uk/software/termine/>.

a documentação, mas não ao projeto e ao *software* para *download* – Apache Stanbol, Ont-O-Mat, FreeCite foi substituído por AnyStyle, RepoMMan, Sherpa/Romeu passou para o projeto Sherpa na versão 2, Simple Automatic Metadata Generation Interface (SamGI), Yahoo Content Analysis API e o *plug-in* do Firefox Dublin Core Viewer Extension. Park e Brenza (2015) discorrem que apesar das ferramentas de GSAM oferecerem muitos benefícios, especialmente no que se refere à racionalização do processo de criação de metadados, existem barreiras significativas à adoção e implementação generalizadas delas. Um fator é que muitas são desenvolvidas localmente para atender às necessidades específicas de um determinado projeto ou como parte da pesquisa acadêmica. Esse ambiente altamente focado para um contexto específico significa que a aplicabilidade geral das ferramentas é potencialmente diminuída. Outro fator, discorrido pelos autores, é o alto nível de conhecimento técnico exigido para seu desenvolvimento e sua incorporação aos fluxos de trabalho diário no processo de criação de metadados. Por fim, ferramentas de metadados semiautomáticas não são testadas em cenários do mundo real, tais como: pequenos tamanhos de amostra, escopo restrito de domínios de projeto e experimentos que não têm objetividade, conforme relatado por Polfreman, Broughton e Wilson (2008).

A pesquisa de Rafferty e demais autores (2015) apresenta um mecanismo de geração de metadados a partir de análises de cliques de vídeo, sendo que esses metadados devem ser usados no suporte ao fornecimento de instruções dinâmicas dentro de um paradigma Smart Home. Os autores utilizaram um método de anotação capaz de gerar metadados enriquecidos para vídeos, sendo criado e implementado dentro de uma plataforma de avaliação chamada Audio BaSEd Instruction ProfiLer (Abseil). Essa plataforma destina-se a trabalhar em conjunto com o repositório de vídeo gerado pelo projeto Personal IADL Assistant (PIA). O objetivo do projeto PIA é ajudar os idosos, oferecendo orientação com atividades instrumentais da vida diária, tais como: preparação de refeições, como utilizar o controle remoto de uma TV, como se barbear, limpar e manter uma casa etc.

Rodrigo e demais autores (2018) apresentam um estudo sobre o ScienceSearch, uma infraestrutura de pesquisa escalável generalizada que utiliza o aprendizado de máquina para capturar metadados de dados, contexto

e artefatos circundantes. A implementação se concentrou no conjunto de dados do Centro Nacional de Microscopia Eletrônica, unidade do Departamento de Energia do Laboratório Nacional Lawrence Berkeley. Esses dados possuem milhões de micrografias produzidas por centenas de cientistas. A problemática identificada foi que os arquivos de micrografia gerados raramente incluem metadados além das configurações de captura do microscópio – por exemplo: exposição, contraste, tensão do sinal. A avaliação de desempenho mostrou que o ScienceSearch é capaz de executar consultas simples em um único nó, em mais de 11 milhões de *tags* de metadados em menos de cinco segundos.

Os estudos de Yang e Park (2018) têm como objetivo apresentar um mecanismo de extração automática para atenuar problemas relacionados à aplicação inconsistente de metadados e à interoperabilidade semântica entre as coleções digitais. Eles sugerem a construção de gráficos conceituais, pois eles têm um bom potencial para facilitar a interpretação adequada dos conceitos de metadados e o uso preciso e consistente dos elementos de dados. Os autores demonstram um mecanismo de extração automática para coleções de bibliotecas chamado ExMETA, que foi projetado utilizando gráficos conceituais como representação interna. A ferramenta é capaz de analisar sentenças em linguagem natural e gerar metadados descritivos e estruturais, permitindo a eliminação de intervenções humanas – ou seja, catalogadoras – que geralmente causam a atribuição incorreta e o mapeamento inconsistente de metadados no processo de produção padrão, como o DC, a partir de dados brutos de coleções digitais.

A pesquisa de Audichya e Asini (2019) teve como objetivo estruturar e padronizar adequadamente o conhecimento disperso sobre a prosódia, denominada de Hindi Poetries, disponível de maneira deficiente ou contraditória em diferentes fontes de informação. Foi utilizada a técnica de linguística computacional e o trabalho de pesquisa também se concentrou em moldar um conjunto de regras padronizadas para a GAM. Os autores testaram um gerador de metadados em 3.026 entradas que incluem diferentes poemas e parte de poemas que cobriam mais de 30 *Chhands*. O resultado do trabalho foi suficiente para provar a robustez da metodologia e do mecanismo técnico do gerador de metadados, que alcançou 98,09% de taxa de

sucesso, juntamente com 1,91% de falha devido a erros de formatação e uso irregular de delimitadores.

Morris (2020) investigou se os códigos de idioma podem ser atribuídos automaticamente aos registros do MARC e avaliar a precisão de qualquer método viável de fazê-lo. A autora enfatiza que nos registros bibliográficos do MARC21, o conteúdo de idioma de um recurso de informação é registrado nas posições do campo 008 35-37, usando um código de três posições de uma lista controlada. Entretanto, uma análise do catálogo da British Library em outubro de 2018 revelou que esse código de idioma não era preenchido em quase 4,7 milhões de registros. Destes, 78% também não possuíam um código para o local de publicação no campo 008 posições 15-17.

Possibilidades de aplicação das ferramentas

Serão demonstradas algumas possibilidades e limitações no uso das ferramentas para GSAM no que tange à qualidade dos metadados e ao desempenho em suas extrações. Para tanto, cada ferramenta mencionada nos artigos anteriormente apresentados foi analisada, partindo-se dos comentários identificados nos artigos encontrados e, quando disponível, nos *sites* das ferramentas utilizadas e mencionados nesses artigos.

Principais possibilidades de aplicação das ferramentas identificadas:

- executar a produção de metadados para conteúdo de aprendizagem com o objetivo de facilitar sua busca e recuperação nos acervos digitais, além de permitir a interoperabilidade (Maratea; Petrosino; Manzo, 2012);
- resolver problemas semânticos nos repositórios, por meio do uso das ferramentas para GAM, propiciando a execução de seu enriquecimento semântico (Vlachidis *et al.*, 2013);
- fornecer informações avançadas para personalização de um sistema para o cliente, ampliando sua capacidade de uso e melhorando a experiência do usuário (Sah; Wade, 2012);
- disponibilizar sistema de catalogação espacial e ferramenta de GAM geoespaciais para tomada de decisão e compartilhamento de dados

entre instituições (Costa *et al.*, 2013). A solução possui possibilidade de ser aplicada em diversas áreas, tais como: previsão do tempo para a agricultura, desastres naturais, desmatamento florestal, defesa e segurança pública, inteligência, comunicação, aquecimento e efeito estufa, saúde, monitoramento dos recursos hídricos e ambientais, tráfego urbano, aéreo e rodovias etc.;

- melhorar o significado dos metadados de objetos vinculados em um repositório por meio da anotação multimídia, aperfeiçoando a recuperação da informação (Verborgh *et al.*, 2012);
- fornecer instruções dinâmicas a partir de clipes de vídeo, facilitando a compreensão e acessibilidade à informação do usuário ao recurso de informação (Rafferty *et al.*, 2015);
- converter dados, documentos e arquivos; extração de metadados de textos e arquivos de imagens; aplicação de metadados em ontologias; análise de citações; criação de metadados para coleções; indexação de documentos etc. (Park; Brenza, 2015);
- gerar e incluir metadados ausentes em repositórios com grandes volumes de dados, objetivando a completude dos dados armazenados (Morris, 2020);
- auxiliar na atenuação de problemas de interoperabilidade semântica entre as coleções digitais (Yang; Park, 2018);
- propor e sugerir metadados de escritos antigos e apoiar sua catalogação, classificação e indexação, objetivando a preservação histórica e cultural (Audichya; Asini, 2019);
- atribuir automaticamente os metadados em um recurso de informação e nos repositórios digitais (Kovacevic *et al.*, 2011).

As limitações encontradas foram:

- o alto grau de especialização dos algoritmos para extração de metadados faz com que eles desconheçam o contexto em que operam, inclusive os que contêm informações valiosas e muitas vezes necessárias.

É necessário aperfeiçoar o tratamento de informações imperfeitas, tais como a incerteza e a incompletude (Verborgh *et al.*, 2012);

- a mineração automática de metadados cognitivos é um desafio, uma vez que é muito difícil compreender automaticamente o conhecimento intelectual subjacente sobre o documento (Sah; Wade, 2012);
- desafio em gerenciar grandes volumes de dados espaciais heterogêneos, assim como melhorar sua organização, busca e prevenir duplicidades (Costa *et al.*, 2013);
- as ferramentas são desenvolvidas localmente para atender as necessidades específicas, altamente focadas para um contexto particular. Além disso, exigem alto grau de habilidade técnica para sua implementação (Park; Brenza, 2015);
- limitação à descrição correta e completa de *metatags*, registros DC e MARC para extração e transformação (Morris, 2020);
- dependência da seleção do vocabulário controlado, da qualidade dos campos de metadados extraídos de *strings* de referência e de um conjunto bem estruturado dos documentos PDF (Maratea; Petrosino; Manzo, 2012);
- escalabilidade e desempenhos impactados por variáveis externas: *hardware*, *software*, sistema operacional, ambiente/plataforma computacional, funcionalidades na própria ferramenta ativadas ou não (Rodrigo *et al.*, 2018);
- indicação de que, em vários casos, os metadados extraídos automaticamente não podem ser inseridos na base de dados, pois dependem de controle da curadoria (Kovacevic *et al.*, 2011);
- ausência de geração de metaconhecimento ou mapa de conhecimento estruturado, como forma de explicitação do conhecimento por meio da documentação, impactando o repasse e o compartilhamento de uso da ferramenta (Yang; Park, 2018);
- necessidade de avaliação da ferramenta em larga escala para analisar o desempenho de extrações de metadados, considerando ambiguidades lexicais e a precisão de anotação do termo (Vlachidis *et al.*, 2013);

- ferramenta depende totalmente da entrada correta e completa dos dados, de acordo com as regras gramaticais, tais como o uso de acentos, de caracteres comuns e especiais, além dos sinais de pontuação (Audichya; Saini, 2019).

Considerações finais

Desse modo, conclui-se que as ferramentas de GSAM descritas na síntese possuem diversas possibilidades de aplicação, das quais citamos: produção de metadados para conteúdos de aprendizagem, resolução de problemas semânticos, fornecimento de informações avançadas para personalização de um sistema para o cliente, disponibilização de metadados geoespaciais para tomada de decisão, melhorar o significado dos metadados de objetos vinculados em uma biblioteca digital, fornecer instruções dinâmicas a partir de clipes de vídeo, geração e inclusão de metadados faltantes em repositórios com grandes volumes de dados, auxiliar na atenuação de problemas de interoperabilidade semântica, gerar metadados de escritos antigos e apoiar a catalogação, a atribuição automática de registros em um recurso de informação, entre outras.

Os gestores de repositório digitais são responsáveis por uma quantidade enorme de metadados relacionados a diferentes tipos de documentos que geralmente são indexados por títulos, assuntos e descritores para que possam ser recuperados posteriormente. Entretanto, nem todos os usuários desses sistemas e repositórios digitais executam a entrada correta e completa de metadados, o que dificulta a recuperação do objeto de pesquisa. O processo de indexação manual de documentos é um trabalho árduo, oneroso e dependendo do volume de dados é humanamente impossível de ser realizado. Essa problemática pode ser mitigada com o desenvolvimento ou a adequação de ferramentas tecnológicas que utilizem técnicas de aprendizagem de máquina para realizar a geração automática dos metadados.

Há lacuna existente na literatura acadêmica de criação de um modelo de referência para geração de metadados para uso geral. Observou-se na pesquisa que as ferramentas de GSAM são desenvolvidas para atender necessidades muito específicas.

Sugere-se pesquisas que mesquem funcionalidades de ferramentas de geração automática e semiautomáticas, com intuito de verificar a complementação das soluções para se atingir um resultado com maior eficiência na geração dos metadados e diminuir o tempo de implementação, auxiliando o processo de catalogação e suporte para os gestores de repositórios digitais em suas tarefas.

Segundo, é importante haver uma ferramenta multilíngue, que suporte diferentes algoritmos de forma flexível e adaptável a diferentes situações e que seja possível o uso de qualquer vocabulário controlado.

Por último, sugere-se testar diversos algoritmos separados e em conjunto para treinamento de um modelo de aprendizagem de máquina, com intuito de obter resultados satisfatórios que cada um pode oferecer no processo de GAM.

Referências

AUDICHYA, M. K.; SAINI J. R. Computational linguistic prosody rule-based unified technique for automatic metadata generation for Hindi poetry. *In: INTERNATIONAL CONFERENCE ON ADVANCES IN INFORMATION TECHNOLOGY*, 1., 2019, Chikmagalur. *Proceedings* [...]. Chikmagalur: [s. n.], 2019.

COSTA, A. *et al.* EURAC SDI: A Near Real Time and Offline Automatic Metadata Generation Processing Chain. *In: GI FORUM*, 2013, Berlin. *Proceedings* [...]. Berlin: [s. n.], 2013. v. 1. DOI 10.1553/giscience2013. Disponível em: https://www.austriaca.at/0xc1aa500d_0x002e6c5e. Acesso em: 15 mar. 2023.

CRYSTAL, A.; LAND, P. Metadata and Search: Global Corporate Circle DCMI 2003 Workshop. *DublinCore*, Washington, Sept. 2003. Disponível em: <http://www.dublincore.org/groups/corporate/Seattle/>. Acesso em: 24 jul. 2020.

CUNHA, M. B. da; CAVALCANTI, C. R. de O. *Dicionário de biblioteconomia e arquivologia*. Brasília, DF: Briquet de Lemos, 2008.

GALVÃO, M. C. B.; RICARTE, I. L. M. Revisão Sistemática da Literatura: conceituação, produção e publicação. *Logeion: Filosofia da Informação*, Rio de Janeiro, v. 6, n. 1, p. 57-73, 2019. DOI 10.21728/logeion.2019v6n1.

p57-73. Disponível em: <https://sites.usp.br/dms/wp-content/uploads/sites/575/2019/12/Revis%C3%A3o-Sistem%C3%A1tica-de-Literatura.pdf>. Acesso em: 15 mar. 2023.

GREENBERG, J. Metadata Extraction and Harvesting: a comparison of two automatic metadata generation applications. *Journal of Internet Cataloging*, [Binghamton], v. 6, n. 4, p. 59-82, 2004.

KITCHENHAM, B. *Procedures for Performing Systematic Reviews*. United Kingdom: Software Engineering Group, 2004. Disponível em: <https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>. Acesso em: 15 mar. 2023.

KOVACEVIC, A. *et al.* Automatic extraction of metadata from scientific publications for CRIS systems. *Program: Electronic Library and Information Systems*, [s. l.], v. 45, n. 4, p. 376-396, 2011.

MARATEA, A.; PETROSINO, A.; MANZO, M. Automatic Generation of SCORM Compliant Metadata for Portable Document Format Files. In: INTERNATIONAL CONFERENCE ON COMPUTER SYSTEMS AND TECHNOLOGIES, 13., 2012, Ruse. *Proceedings* [...]. Ruse: [s. n.], 2012. p. 360-367.

MEY, E. S. A.; SILVEIRA, N. C. Considerações teóricas aligeiradas sobre a catalogação e sua aplicação. *InCID: Revista de Ciência da Informação e Documentação*, Ribeirão Preto, v. 1, n.1, p. 125-137, 2010.

MOHER, D. *et al.* Preferred Reporting Items for Systematic Reviews and MetaAnalyses: The PRISMA Statement. *PLoS Medicine*, [s. l.], v. 6, n. 7, p. 1-6, July 2009. DOI 10.1371/journal.pmed1000097. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2707599/pdf/pmed.1000097.pdf>. Acesso em: 12 nov. 2023.

MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. *American Documentation*, Washington, v. 2, n. 1, p. 20-32, Jan. 1951.

MORRIS, V. Automated Language Identification of Bibliographic Resources. *Cataloging & Classification Quarterly*, [s. l.], v. 58, n. 1, p. 1-27, 2020.

PARK, J.-R.; BRENZA, A. Evaluation of Semi-Automatic Metadata Generation Tools: A Survey of the Current State of the Art. *Information Technology and Libraries*, Chicago, v. 34, ed. 3, p. 22-42,

- Sept. 2015. Disponível em: <https://www.proquest.com/docview/1722657915?sourcetype=Scholarly%20Journals>. Acesso em: 15 mar. 2023.
- POLFREMAN, M.; BROUGHTON, V; WILSON, A. Metadata Generation for Resource Discovery. *Textarchive.ru*, [s. l.], 2008. Disponível em: <https://textarchive.ru/c-2308841-pall.html>. Acesso em: 15 mar. 2023.
- RAFFERTY, J. *et al.* Automatic Metadata Generation Through Analysis of Narration Within Instructional Videos. *Journal of Medical Systems*, [s. l.], v. 39, n. 94, p. 92-95, 2015. DOI 10.1007/s10916-015-0295-2. Disponível em: <https://link.springer.com/article/10.1007/s10916-015-0295-2>. Acesso em: 15 mar. 2023.
- REINSEL, D.; GANTZ, J.; RYDNING, J. *The Digitization of the World: From Edge to Core*. [S. l.]: IDC White Paper, 2018. Disponível em: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>. Acesso em: 15 mar. 2023.
- RODRIGO, G. P. *et al.* ScienceSearch: Enabling Search through Automatic Metadata Generation. In: INTERNATIONAL CONFERENCE ON E-SCIENCE, 14., 2018, United States. *Proceedings [...]*. United States: [s. n.], 2018. DOI 10.1109/escience.2018.00025. Disponível em: <https://escholarship.org/content/qt7t14b5h1/qt7t14b5h1.pdf>. Acesso em: 15 mar. 2023.
- SAH, M.; WADE, V. Automatic metadata mining from multilingual enterprise content. *Journal Web Semantics*, [s. l.], v. 11, p. 41-62, Mar. 2012.
- VERBORGH, R. *et al.* Enabling context-aware multimedia annotation by a novel generic semantic problem-solving platform. *Multimed Tools Appl*, [s. l.], n. 61, p. 105-129, 2012.
- VLACHIDIS, A. *et al.* Automatic Metadata Generation in an Archaeological Digital Library: Semantic Annotation of Grey Literature. In: PRZEPIÓRKOWSKI, A. *et al.* (ed.). *Computational Linguistics: Studies in Computational Intelligence*. Berlin: Springer, 2013. v. 458.
- YANG, G.; PARK, J. Automatic Extraction of Metadata Information for Library Collections. *International Journal of Advanced Culture Technology*, [s. l.], v. 6, n. 2, p. 117-122, 2018.

CAPÍTULO 11

Comunicação científica no Brasil e a Plataforma Lattes

Marcos Luiz Mucheroni

Valmir Batista Prestes e Souza

Introdução

As tecnologias estão presentes na sociedade há muito tempo e a cada momento o homem recria o modo de interagir e viver em sociedade, sendo que essa transformação traz consigo mudanças significativas no modo de conhecer e reconhecer os recursos e os dispositivos existentes ou advindos dessas mudanças.

Tanto o desenvolvimento inicial da internet – na metade da década de 1970 – como o desenvolvimento da *web* – início dos anos 1990 – tinham como grande motivação inicial uma mais rápida, segura e precisa divulgação da informação científica, por isso ela se desenvolveu primeiro nos meios acadêmicos.

Como a ciência busca desvendar e compreender a natureza, a sociedade – tanto material quanto a imaterial – e seus fenômenos, a forma de tornar os métodos sistemáticos e seguros é passar por uma ampla revisão pela comunidade científica indo até a publicação dos resultados, em que há necessidade da chamada revisão por pares. Nesse viés, a ciência é uma instituição social, dinâmica, contínua e cumulativa e “sem pretensões históricas”, que se insere e influencia a sociedade (Targino, 2000, p. 2).

Dessa forma, o avanço da ciência e o uso de dispositivos informacionais contemporâneos proporcionam mudanças na produção e comunicação dos resultados de pesquisas que têm incorporado as tecnologias emergentes, pois o pesquisador necessita de informação para apropriar-se de conceitos e teorias desenvolvidas pela comunidade científica e desenvolver

experimentos que visam dirimir as incertezas decorrentes da dialética que aborda o estudo.

As pesquisas realizadas e acessíveis são subsídios na construção de concepções que corroboram ou refutam os resultados anteriormente publicados. No primeiro momento, a busca por informação consiste no insumo necessário para fundamentar e orientar os desencadeamentos teóricos sobre o tema requerido, ou seja, a informação publicada nos canais de comunicação consiste no aporte teórico que subsidiará o pesquisador.

Posteriormente, o pesquisador passará de usuário para produtor de informação e essa transformação ocorre após avaliação da comunidade científica que analisa se os critérios explicitados no trabalho e o caráter científico foram alcançados no estudo e, sendo aceito, o resultado da pesquisa será publicado e disponibilizado, ocorrendo assim a produção e a comunicação científica.

Há um aumento significativo de literatura e o cientista busca identificar os resultados de pesquisas veiculados nos meios de comunicação que de alguma forma colaboram nos estudos que estão desenvolvendo ou que irão desenvolver. Há uma variedade de periódicos *on-line*, uns reconhecidos pela comunidade científica e tantos outros que estão se tornando conhecidos e muitas vezes passam despercebidos.

Portanto, esta produção e comunicação proporcionam, além do acesso às informações sobre pesquisas realizadas, um canal formal que avalia os resultados submetidos pelos pesquisadores para posterior disponibilização à comunidade.

Dessa forma, a comunicação científica vem passando por mudanças quanto à inserção das tecnologias digitais, abrangência e visibilidade dos resultados de pesquisa, que permite acesso à informação produzida pela comunidade.

Essa mudança contribui com a ciência aberta e o desenvolvimento de padrões de interoperabilidade para intercambiar os dados existentes e promover o acesso à informação das pesquisas produzidas.

Ressalta-se que esse cenário não pode ser reduzido apenas à inserção de novas tecnologias, à abrangência ou à visibilidade da ciência, mas ao

contexto no qual ela está inserida quanto a mudanças culturais de acesso inclusivo na igualdade de oportunidade e a mudanças de políticas.

Embora os aspectos de relevância, precisão e veracidade de informação na *web* possam e devam ser questionados, as mídias sociais e a difusão de informação pela via digital já mudaram profundamente os aspectos da comunicação científica atual.

Comunicação científica e informação

Compreender o processo de comunicação científica permite visualizar o fluxo informacional e a política nos quais os pesquisadores estão inseridos, organizados por áreas do conhecimento e a relação das informações entre as áreas.

A realização da pesquisa e sua comunicação junto à comunidade científica são inseparáveis dos resultados, e isto também é feito por relatos orais, em seminários, até a conclusão final e a obtenção segura dos resultados, que devem ser divulgados ao grande público.

O registro informal e formal contribui com a disseminação do resultado da pesquisa científica, proporciona reflexão sobre o objeto de estudo, além de auxiliar na construção de *networking* de pesquisadores em sua área de atuação, ocorrendo a aceitação e validação do estudo pelos pares, ou mesmo a desaprovação por ausência de procedimentos ou dados científicos.

Nesse aspecto, pode-se observar que os pesquisadores buscam diminuir a distância existente entre os usuários que utilizam e produzem a informação científica, embora esse movimento seja gradual e contínuo, este perpassa por décadas de esforços. Sendo assim, Marcondes e Sayão (2002) destacam os aspectos de dispersão espacial e temporal que envolvem o processo de produção e comunicação dos resultados, plataformas padronizadas de pesquisadores ajudam o processo.

Podemos descrever que a *dispersão espacial* pode ser caracterizada pelo distanciamento físico/geográfico: 1) entre pesquisadores que compartilham os anseios de aprofundar os estudos em determinado assunto, mas têm que contornar a distância dos centros de pesquisas, grupos de pesquisas, pesquisadores e demais atores que influenciam ou são influenciados pelo objeto de estudo; 2) entre a produção de ciência e sua publicização, que se

encontra dispersa. Embora convirjam nos anseios de aprofundar os estudos no assunto, ocorrem em ambientes que necessitam interoperar.

Já a *dispersão temporal* pode ser caracterizada pelo distanciamento do período em que ocorre a pesquisa e sua publicação: 1) seja pelo tempo de realização, coleta, processamento e escrita dos resultados até sua comunicação; 2) seja pela continuidade da pesquisa e/ou sua replicação por outros cientistas.

Os fatores de infraestrutura e o processo de avaliação devem ser compreendidos pelo pesquisador e constituem um processo intrínseco à pesquisa desde sua concepção, durante todas as etapas do desenvolvimento, e se transformam com o passar do tempo para melhorar a fluidez das informações.

Meadows (1999) descreve que a comunicação científica vem passando por mudanças aceleradas com o avanço da tecnologia da informação. Novos meios de comunicação estão sendo utilizados para atender às necessidades da comunidade científica em constante evolução e o ponto básico continua a ser as características do veículo/meio de comunicação a ser utilizado.

Targino (2000) afirma que a comunicação faz parte do ser humano e esta perpassa pelos achados da ciência, em que o emissor e o receptor compartilham os signos e significados necessários para manutenção do diálogo, no qual obedece às práticas estabelecidas pela comunidade científica. Logo, o processo de comunicação pressupõe um estoque comum de elementos preexistentes – linguagem, expressões, códigos etc. –, essencial para facilitar o fluxo informacional, no qual o cientista busca literatura “exclusiva” às demandas singulares.

Assim, as práticas da comunidade científica e tecnológica foram alteradas e a segmentação em função das especialidades é fator de fragmentação e imprecisão científica.

Então, os resultados da ciência são tão importantes quanto as etapas utilizadas para elaboração deles, com apresentação de materiais e métodos que permitam sua reprodução, fazendo com que outros pesquisadores possam repetir os experimentos.

Conforme já foi salientado, a comunicação científica auxilia a sedimentação e o avanço das áreas. No entanto, aspectos de complexidade e continuidade que se entrelaçam com campos vizinhos do conhecimento são a prova de que se não houver comunicação eficiente, os avanços podem ser impedidos.

No Brasil, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) promove e incentiva que os registros de ações realizadas pelos pesquisadores nas instituições de pesquisa sejam inseridos na Plataforma Lattes, sendo aceitos por instituições de fomento à pesquisa e outras instituições como referência para identificar o *know-how* do pesquisador quanto às atividades desenvolvidas.

Diante disso, destaca-se que o processo de registro de informações na Plataforma Lattes promove a divulgação das atividades realizadas pelos pesquisadores, pois permite recuperá-las ao tempo em que lança desafios internos na organização para manusear a informação em seus processos informacionais, tanto para consumo interno voltado a usos administrativos, ao planejamento e a políticas de desenvolvimento e inovação, quanto para a publicização das pesquisas realizadas.

A Plataforma Lattes

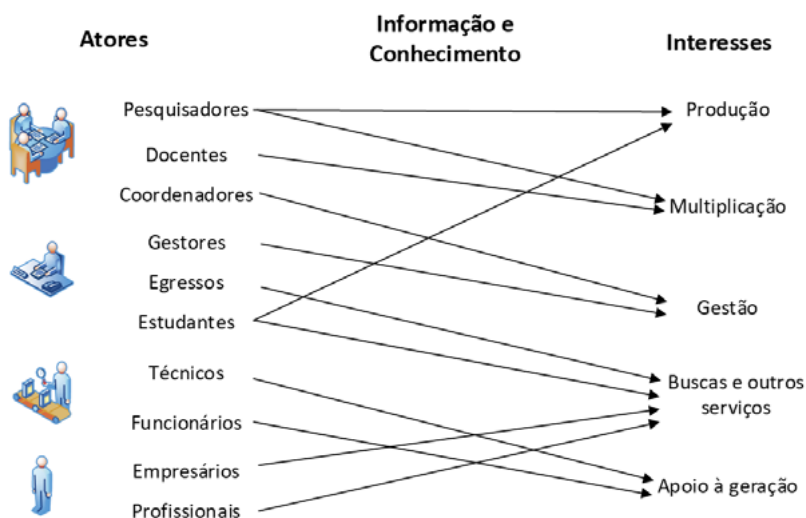
A dispersão espacial e temporal que Marcondes e Sayão (2002) apresentam tanto existe internamente na comunicação científica em áreas bem estabelecidas de estudo, como na sua interdisciplinaridade e transdisciplinaridade em outras áreas. Boa parte dos métodos e avanços epistemológicos estão limitados aos contornos da área, configurando, assim, o estranhamento com as tecnologias, por exemplo, como fruto de um certo fechamento, já que ela pode ajudar na abertura e desenvolvimento inter e entre áreas que estão relacionadas, assim como maior precisão temporal e espacial das pesquisas e auxiliar até mesmo na coleta dos dados.

Por isso, estudos envolvendo ferramentas tecnológicas são necessários e importantes neste novo cenário, e com isso, o uso de currículos e suas plataformas que atestem a produção científica e deem visibilidade a ela são fundamentais.

A necessidade de parâmetros uniformes para acompanhar e promover o desenvolvimento de pesquisa no Brasil levou o CNPq a sistematizar a iniciativa conhecida e difundida como Plataforma Lattes para agregar informações utilizadas por diversos atores no contexto da Ciência, Tecnologia e Inovação (CT&I) do país.

Além de pesquisadores e agências de fomento, há outros interessados no acompanhamento e desenvolvimento de CT&I, e possuem interesses diversos, mas com pontos em comum na produção científica e tecnológica nacional (Figura 1).

Figura 1 - Atores no processo de informação e conhecimento



Fonte: adaptada de Pacheco e Kern (2003).

A interação dos atores (Figura 1) no desenvolvimento de pesquisas caracteriza a dimensão de informações que circundam a dinâmica institucional de promoção da pesquisa regional e nacional que, nas últimas décadas, vem desenvolvendo base de dados para facilitar a organização e recuperação no sistema informacional, permitindo explorar os registros sobre resultados de pesquisas.

Nesse sentido, Targino (2000) descreve que o pesquisador busca por literaturas “exclusivas” de seus interesses e por informações pertinentes às suas demandas mais singulares. Uma fonte de informação que vem sendo demandada para mensurar a ciência produzida é a Plataforma Lattes do CNPq.

Segundo o CNPq (2014a), a Plataforma Lattes corresponde à integração das bases de dados do currículo profissional, grupos de pesquisa e instituições em um único sistema de informação, que possibilita a criação de políticas

públicas do Ministério de Ciência e Tecnologia e de outros órgãos governamentais.

A Plataforma Lattes ou base de dados Lattes concentra os dados de pesquisadores, grupos de pesquisa e instituições que movem as engrenagens da ciência brasileira; Lane (2010) descreve que essa iniciativa corrobora com a infraestrutura de dados confiável, aberta e representa uma boa prática.

Vale destacar que o CNPq, ao desenvolver o sistema de informação, promoveu a uniformização no currículo do pesquisador, que era tido como disperso e dificultava a identificação de informações sobre as ações que estavam sendo realizadas, bem como *know-how* dos pesquisadores que buscam por investimentos para conseguir fazer pesquisa no país.

É notória a importância desse sistema de informação que desde sua implantação vem sendo utilizado por uma necessidade represada de identificar os atores envolvidos no desenvolvimento de pesquisa dentro da abrangência do território nacional. Além disso, nota-se a importância de políticas públicas voltada à promoção e ao fomento no contexto da CT&I.

Kern (2010) descreve a Plataforma Lattes como um sistema socio-tecnológico que depende da colaboração dinâmica entre os agentes humanos e artificiais e os objetos em torno dos quais colaboram – dessa forma, todas as ligações desses objetos são relevantes para a análise sistêmica –, o qual segue a definição de Bunge (2014), que apresenta que qualquer sistema concreto pode ser definido pelo modelo Composition, Environment, Structure e Mechanism (CESM)¹: componentes (itens do ambiente), estrutura (ligações entre componentes e destes com itens do ambiente) e mecanismos (processos que geram a novidade qualitativa), conforme pode ser visualizado no Quadro 1, a seguir.

1 Segundo Bunge (2014), qualquer sistema pode ser modelado, em qualquer dado instante, como o quádruplo CESM:

$\mu(s) = \langle C(s), E(s), S(s), M(s) \rangle$, onde

$C(s)$ = Composição: coleção de todas as partes de s ;

$E(s)$ = Ambiente: conjunto de itens, exceto aqueles em s , que agem ou são influenciados por alguns ou todos os componentes de s ;

$S(s)$ = Estrutura: conjunto de relações, em particular vínculos, entre componentes de s ou entre estes e itens em seu ambiente $E(s)$.

$M(s)$ = Mecanismo: conjunto de processos em s que o fazem se comportar da maneira que o faz.

Quadro 1 – Quadro-síntese do modelo CESM da Plataforma Lattes sociotecnológica

Sistema	Composição	Ambiente	Estrutura	Mecanismo
Plataforma Lattes Sociotecnológica	Pesquisadores e professores; estudantes e egressos; especialistas e profissionais; <i>staff</i> do CNPq; agentes de Extração, Transformação e Carregamento (ETL); agentes de análise e síntese; agentes conectores; currículos.	Organizações de C&T; grupo de pesquisa; agências governamentais; empresários e empresas; jornalistas e mídia; fontes de informação externas; infraestrutura de Tecnologia da Informação e Comunicação (TIC); a <i>web</i> ; cultura CT&I.	Atualização curricular; busca de informação; monitoramento e denúncia; ETL; vínculos interpessoais; vínculos institucionais; <i>links</i> , inclusive mútuos; Análise e síntese de informação; construção de consenso; publicação e acesso a currículos; interesses informacionais; intercâmbio com fontes externas; acesso via TIC; influência cultural.	Apropriação individual do currículo; transparência e autorregulação; esquema curricular por consenso; institucionalização do Curriculum Vitae (CV); sinergia com outros projetos; currículo mandatário no fomento.

Fonte: Kern (2010, p. 61).

Dentre os recursos da Plataforma Lattes, há de se destacar o armazenamento das informações inseridas pelos pesquisadores referente às atividades laborais realizadas junto às instituições de pesquisas e estão dispostas no Sistema de Currículo Lattes, desenvolvido na década de 1990.

Kern (2010) prossegue ao apresentar que as *agências de fomento* demandam informação curricular nos processos de fomento à CT&I e distribuem recursos sob influência dessas informações. Os *empresários e empresas* são potenciais interessados em cooperação com os detentores do conhecimento científico-tecnológico. *Jornalistas e mídia* encontram na plataforma um registro adequado para a seleção de fontes jornalísticas. Além desses atores, há também a ação dos pesquisadores e organizações de CT&I.

Essa dinâmica entre os atores está representada na Figura 1, na qual a busca por informação científica e sua manutenção compõem o escopo do sistema de informação do CNPq, cujos interesses são voltados à produção, multiplicação, gestão, ao apoio à geração e à busca por informação.

Ainda que a Plataforma Lattes esteja em constante melhoria, ela se destaca por esses interesses, além de procurar observar coerências dos dados – Digital Object Identifier (DOI), Journal Citation Reports (JCR) etc. – e atualização destes.

Currículo Lattes

O Currículo Lattes concentra as informações de atividades relacionadas às diversas ações em que o pesquisador pode desenvolver durante o percurso formativo e profissional, dentre elas destacam-se a *formação acadêmica* – graduação, especialização, mestrado, doutorado –, a *atuação profissional*, os *prêmios e títulos*, a *produção bibliográfica* – trabalhos em eventos, artigos publicados, livros e capítulos, texto em jornais e revistas –, a *produção técnica* – *software* e produto tecnológico –, *projetos de pesquisa* e *patentes*.

O registro das informações pelo pesquisador e sua manutenção por longo período no sistema de informação possibilitam identificar as atividades realizadas na vida pregressa e as atividades que estão sendo desenvolvidas.

O sistema permite consultar pelo nome do pesquisador ou pelo assunto – título ou palavra-chave da produção –, tendo como resultado a lista de currículos de pesquisadores que atendem aos parâmetros fornecidos na busca, podendo aplicar filtros para refinar a busca e o resultado sempre será o currículo do pesquisador.

Destacamos que, com a informação contida nessa base de dados, é possível ir além do currículo do pesquisador, mas atualmente o sistema está projetado apenas para gerar a lista de currículos, não sendo possível identificar e/ou gerar uma lista de artigos publicados sobre um determinado tema e/ou assunto ou mesmo a relação de uma publicação científica com atividade de ensino ou atividade técnica, visto que o Currículo Lattes foi concebido para registrar a atividade desenvolvida pelo pesquisador e o retorno da consulta é disposto no formato de currículo profissional. Embora seja amplo, ele permite solicitar os dados em formatos mais resumidos.

Mesmo considerando que a importância do sistema como fonte de informação esteja registrada em vários trabalhos, vale destacar que o acesso ao conjunto de dados restringe muito o desenvolvimento de pesquisas e ações

internas das instituições, pois requerem ferramentas para coleta e análise dos dados, como, por exemplo, Lattes Extrator², scriptLattes³, Sucuripa, Somos⁴ e Stela Experta⁵, conforme o Quadro 2.

Um conjunto de ferramentas permite ainda extrair, alterar e descrever dados do Lattes, feito em formato de código livre, algumas solicitações exigem contato com o administrador.

Quadro 2 - Ferramentas que acessam a Plataforma Lattes

Recurso	Público	Objetivo
Lattes Extrator	Instituições cadastradas na Plataforma Lattes – instituição CT&I; universidades; agências de fomento à pesquisa.	Permite coletar os dados das atividades dos pesquisadores da instituição para acompanhar a produção intelectual.
scriptLattes	<i>Software</i> livre de código aberto e sob a licença GNU-GPL, não há restrição para utilizá-lo.	Desenvolvido para a extração e compilação automática de: (1) produções bibliográficas, (2) produções técnicas, (3) produções artísticas, (4) orientações, (5) projetos de pesquisa, (6) prêmios e títulos, (7) grafo de colaborações, (8) mapa de geolocalização, e (9) coautoria e internacionalização de um conjunto de pesquisadores cadastrados na Plataforma Lattes (Mena-Chalco; Cesar-Junior, 2024)
Sucuripa	Curso de pós-graduação.	Assegurar a qualidade dos cursos de mestrado e doutorado no país, por meio do acompanhamento das atividades que são realizadas.
Somos	<i>Software</i> não gratuito; instituição de pesquisa que busca disponibilizar suas informações constantes na Plataforma Lattes Empresas e para a sociedade em geral.	Incrementar a interação da universidade em áreas de pesquisa científica e tecnológica com instituições públicas e privadas.
Stela Experta	<i>Software</i> não gratuito; Instituição de Ensino Superior (IES).	Ofertar às IES brasileiras serviços de apoio estratégico nas áreas de gestão de ensino, pesquisa, extensão e inovação.

Fonte: elaborado pelos autores.

Instituições podem utilizar a informação curricular dos pesquisadores em ações internas e de forma isolada: Currículo Lattes como documento

.....

2 Disponível em: <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>.

3 Disponível em: <http://scriptlattes.sourceforge.net/>.

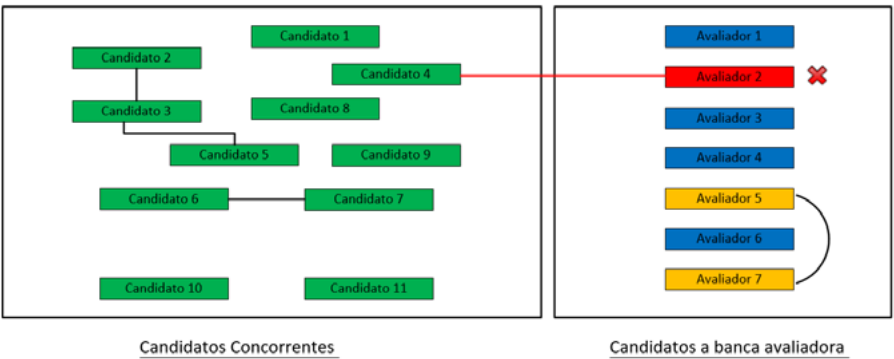
4 Disponível em: <http://www.somos.ufmg.br/sobre>.

5 Disponível em: <https://www.stelaexperta.com.br/unioeste/index.html>.

para seleção de candidatos na carreira de magistério superior, seleção de mestrado e doutorado, concessão de bolsa de iniciação científica e progressão de docentes na carreira.

Nesse contexto, Mena-Chalco (2015) apresenta o método de redes de coautoria – grafos de colaborações – bastante utilizado em trabalhos científicos, assim como a rede de coocorrência para uso em atividade administrativa de indicação de banca examinadora em concursos públicos em que não pode haver vínculo entre o candidato e membros da banca, como exemplo, a Figura 2.

Figura 2 – Informação curricular para indicação de banca de avaliação



Fonte: Mena-Chalco (2015).

A rede de coocorrência de autores permite identificar a relação entre o candidato e o membro a ser indicado para compor a banca examinadora, permitindo assim a avaliação isenta de vínculo acadêmico entre o avaliado e o avaliador. No exemplo, o candidato 4 possui vínculo de colaboração com o avaliador 4, Mena-Chalco (2015) acrescenta que é possível avaliar relações acadêmicas de segunda ordem entre candidatos concorrentes e possíveis membros de banca, em que os avaliadores 5 e 7 possuem vínculo de colaboração entre si, sendo ideal ter membros avaliadores sem nenhuma interação prévia entre eles. Em estudo recente pode-se observar a aplicação da referida abordagem no contexto de uma agência de fomento do Uruguai (Rodrigues; Tomassini-Urti; Mena-Chalco, 2023).

A informação curricular para indicação de banca de avaliação é um exemplo do uso das informações contidas nos Currículos Lattes como forma de melhorar a transparência dos atos administrativos das instituições. Nesse sentido, Silva e demais autores (2012) reforçam a importância das informações para gestão, podendo atuar como instrumento fundamental para as atividades de geração de indicadores de produção científica, possibilitando o tratamento e a difusão das informações necessárias à formulação e à gestão de políticas de IES.

Além disso, uma das atividades de autoavaliação de cursos de graduação e pós-graduação é identificar a publicação de artigos científicos dos docentes e discentes e as atividades do egresso da universidade na sociedade.

Currículo Lattes como fonte de informação

Estudos que analisam o fluxo da produção e comunicação científica no país visando conhecer sua dinâmica traduzem o que Targino (2000) descreve como a busca por literaturas “exclusivas de seus interesses e informações pertinentes às suas demandas mais singulares”. Uma fonte de informação que vem sendo demandada para mensurar a ciência produzida é a Plataforma Lattes.

Logo, o Currículo Lattes constitui uma fonte de informação de valor sobre a ciência brasileira e permite a consulta pública aos currículos dos pesquisadores e sua produção científica, possibilitando identificar os locais selecionados para publicação dos resultados de pesquisas (Mugnaini *et al.*, 2019). Entretanto, Mugnaini, Leite e Leta (2011) informam que a coleta dos dados na base do Lattes demanda tempo do pesquisador, que busca analisar a comunicação científica para estudos cientométricos, por exemplo.

Já os autores Silva e Smit (2009) fazem uma reflexão sobre a Plataforma Lattes na *perspectiva de arquivos pessoais* e na *perspectiva dos procedimentos de preenchimentos do sistema*. Quanto à primeira, a Plataforma Lattes diverge da concepção arquivística, pois, por permitir o registro das atividades individualmente, sem a possibilidade de agregação posterior, limita compreender o contexto e a sequência das atividades que foram realizadas.

Quanto aos *procedimentos de preenchimentos do sistema*, Silva e Smit (2009), categorizam as permissões que o pesquisador possui no sistema para

inserir as atividades realizadas, o que pode implicar negativamente na recuperação da informação, caso não seja observado pelo usuário no momento da inserção:

Autonomia Total: o usuário tem a liberdade de cadastrar as palavras que desejar, sem restrição ou qualquer direcionamento. Ex: Título de uma publicação;

Autonomia Parcial: São campos em que, inicialmente, se tem autonomia total, porém cada novo termo cadastrado pelo usuário é automaticamente armazenado no sistema, que vai criando uma lista de termos exclusiva do usuário. Nas vezes seguintes em que o usuário inserir outros termos, será possível consultar e adotar termos anteriormente criados por ele. O usuário pode excluir qualquer termo dessa lista, ou incluir novos;

Sem Autonomia: o sistema prevê, inicialmente, opções que o usuário deve selecionar. Entretanto, a existência dessas opções prévias não impede que novos termos sejam incluídos, caso o usuário não se satisfaça com as opções oferecidas. Ex.: Áreas do Conhecimento, Setores de aplicação (Silva; Smit, 2009, p. 82).

Dentro desse contexto, as informações na Plataforma Lattes são demandadas constantemente por serem fonte de informação de valor, mas requerem cuidados por quem vai utilizá-las, visto que os dados ali inseridos podem conter falhas oriundas de preenchimento inadequado, além do problema de desatualização por parte dos detentores dos currículos (Digiampietri *et al.*, 2014).

Metadados

Os metadados são utilizados em sistemas computacionais como fonte de informação para identificar e promover a uniformização do conteúdo proveniente de sistemas externos, visando ao intercâmbio de dados ou, quando não é possível, à menos a integração de dados. A biblioteconomia faz uso do metadado para melhorar a capacidade do estoque informacional, visando aperfeiçoar o sistema de recuperação da informação e comunicação.

Dessa forma, o compartilhamento de informação proporciona agilidade em atividades operacionais e de recuperação da informação, desde a inserção até o acesso dos elementos informacionais, de acordo com a disponibilidade do recurso. O desenvolvimento e a compreensão dos aspectos que envolvem o compartilhamento contribuem para ações individuais e ao mesmo tempo com a informação no todo, pois caracteriza-se por um sistema interligado.

O conjunto de dados primários, compartilhado na instituição por um sistema de banco de dados, promove o acesso em tempo real a mudanças ocorridas internamente e registradas individualmente no sistema. Entretanto, a depender do contexto, é necessário obter dados externos, dados secundários, que venham a corroborar com as atividades operacionais e de recuperação da informação.

Dessa maneira, o uso de metadados que descrevem os objetos digitais permite intercâmbio entre quem necessita do conjunto de dados e quem dispõe dele, pois os parâmetros possibilitam a comunicação entre o emissor e o receptor de forma transparente. Além disso, não compromete a operação individual de cada sistema, pois inibe possível falha de integridade, confiabilidade e segurança dos sistemas envolvidos. O desenvolvimento dos metadados para uso em ambientes computacionais busca promover soluções de como organizar e permitir acesso à informação digital e ao mesmo tempo acessar outras, sem impactar nas atividades operacionais da instituição. Sua estrutura é variável desde sua concepção, buscando descrever, identificar, definir, recuperar, preservar e intercambiar o objeto de informação disponível na *web*, sendo que a descrição da informação contida nessa estrutura permite acessar e avaliar a informação disponível, sem obrigatoriamente acessar o documento originário.

Portanto, intercambiar informações entre sistemas heterogêneos, indicando o recurso para acessar a informação, se esta estiver disponível. Logo, os metadados devem possuir informações suficientes para identificação do objeto, permitindo que este seja recuperado.

A partir do acesso ao metadado pode-se identificar a disponibilidade do documento que o originou e avaliar se a informação descrita atende ao acesso à informação e à utilidade dela. Entretanto, para tornar a informação

inequívoca e explorar o objeto digital, de acordo com a área do conhecimento e seu significado, é requerido o uso de identificador uniforme de recurso, do acrônimo URI.

National Information Standards Organization (2006) descreve que os identificadores de recursos permitem que os sistemas de informação intercambiem objetos digitais e ultrapassa o gerenciamento de documentos tradicionais. Berners-Lee concebeu o URI para identificar o recurso de informação esteja ele presente na *web* ou não, entretanto, o seu uso ganha destaque na *web* semântica, pois permite que o recurso digital seja compreendido pelo agente não humano de forma única.

Os metadados podem ser categorizados como descritivos, estruturais, arquivos físicos e administrativos:

Metadados Descritivos: Descreve a entidade intelectual por meio de propriedades como autor e título, e suporta a descoberta e a entrega de conteúdo digital. Pode também fornecer um contexto histórico, por exemplo, especificando qual material baseado em impressão era a fonte original para um derivativo digital (proveniência de origem);

Metadados Estruturais: Captura relações estruturais físicas, como qual imagem é incorporada dentro de qual *site*, bem como relações estruturais lógicas, como qual página segue qual em um livro digitalizado;

Metadados de arquivos físicos: Inclui informações técnicas que se aplicam a qualquer tipo de arquivo, como informações sobre o *software* e o *hardware* em que o objeto digital pode ser renderizado ou executado, ou somas de verificação e assinaturas digitais para garantir a correção e a autenticidade. Ele também inclui informações técnicas específicas do tipo de conteúdo, como a largura da imagem de uma imagem ou o tempo decorrido de um arquivo de áudio;

Metadados Administrativos: Inclui informações de proveniência de quem cuidou do objeto digital e quais ações de preservação foram executadas nele, bem como direitos e informações de permissão que especificam, por exemplo, o acesso ao objeto digital, incluindo quais

ações de preservação são permitidas (Dappert; Enders, 2010, p. 6, tradução nossa).

A abordagem de metadados é importante para compreender que as informações presentes no Currículo Lattes podem ser acessadas por meio de metadados que a própria plataforma disponibiliza, sendo uma fonte de informação importante sobre as ações realizadas pelos pesquisadores e por eles inseridas no Lattes.

A semântica dos dados disposta nos metadados pode ser descrita por intermédio do modelo de dados ou da estrutura de descrição de recurso que fornece vocabulários legíveis tanto por humanos quanto processáveis por máquina, sendo um conjunto de propriedades, ou elementos de metadados, definidos por comunidades de descrição de recursos (Miller, 1998). O Resource Description Framework (RDF) possibilita interoperar metadados mediante o suporte de semântica, sintaxe e estrutura, fornecendo a capacidade para essas comunidades definirem elementos de metadados conforme necessário e utiliza o eXtensible Markup Language (XML) como uma sintaxe comum para a troca e o processamento de metadados.

Pelo fato de a XML ser uma linguagem que permite ao criador elaborar suas próprias *tags*, torna-se possível desenvolver outras linguagens de marcação a partir da XML. Diversos padrões para marcação de documentos podem ser construídos baseados na estrutura que o XML oferece. Já existem dezenas, talvez centenas, de outras linguagens que foram desenvolvidas a partir do XML nas mais diversas áreas do conhecimento. Isso possibilita que comunidades formem vocabulários próprios, fornecendo significado a termos específicos e utilizados em conjunto. A criação de novas linguagens é possível por meio da utilização do Document Type Definition (DTD) ou XML Schema, que é a evolução daquele.

No arquivo XML, os elementos (*tags*) são repetidos quantas vezes for necessário para descrever o objeto numa estrutura aninhada, e isso compõe a flexibilidade e o reuso desse recurso, pois o acréscimo de novos elementos ao XML não inviabiliza ou prejudica o sistema anterior, pois tais elementos são ignorados permitindo a interoperabilidade dos sistemas. Para exemplificar, foi realizada uma “evolução no Currículo Lattes a fim de permitir

o registro dos períodos de licença-maternidade” (Conselho Nacional de Desenvolvimento Científico e Tecnológico, 2024b). Em abril de 2021 essa nova *tag* está disponível nos metadados da plataforma e não impede a integração de dados, caso o sistema receptor deixe de inserir ou interpretar a licença-maternidade disponível no novo arquivo XML.

A integração de dados, ou mesmo o intercâmbio de dados por meio da Linguagem XML, permite uso de recurso adicional no modelo hierárquico – estrutura em árvore – do documento, como a DTD (Sousa; Martins; Ramalho, 2018).

A DTD é opcional no arquivo XML, cuja função é definir as informações e os tipos de informações presentes no documento, sem restringir os tipos básicos do dado, como inteiro e *string*, mas descrever o subelemento e atributo dentro de um elemento (Silberschatz; Korth; Sudarshan, 2006).

Nesse sentido, a figura a seguir apresenta DTD do elemento doutorado, seus atributos e subelementos, disposto no Currículo Lattes, no qual os atributos (linha 3 a 37) do elemento doutorado são alfanuméricos (CDATA), sendo que o DTD define a obrigatoriedade de sequência e disposição destes, por exemplo, subelementos – palavras-chaves, áreas-do-conhecimento e setores-de-atividades – e informa que eles são opcionais e representados pelo operador (símbolo ?), na primeira linha da Figura 3.

Figura 3 – DTD: Elemento Doutorado, atributos e subelementos

```

1 <!ELEMENT DOUTORADO (PALAVRAS-CHAVE?, AREAS-DO-CONHECIMENTO, SETORES-DE-ATIVIDADE?)>
2 <!--
3   SEQUENCIA-FORMACAO CDATA IMPLIED
4   NÍVEL CDADOS IMPLÍCITOS
5   CODIGO-INSTITUICAO CDATA IMPLIED
6   NOME-INSTITUICAO CDATA IMPLIED
7   CODIGO-ORGAO CDATA IMPLIED
8   NOME-ORGAO CDATA IMPLIED
9   CODIGO-CURSO CDATA IMPLIED
10  NOME-CURSO CDATA IMPLIED
11  CODIGO-AREA-CURSO CDATA IMPLIED
12  STATUS-DO-CURSO (EM ANDAMENTO CONCLUIDO INCOMPLETO) IMPLIED
13  CDADOS DE INÍCIO DO ANO IMPLÍCITOS
14  ANO-DE-CONCLUSÃO CDATA IMPLIED
15  FLAG-BOLSA CDADOS IMPLÍCITOS
16  CDADOS DA AGÊNCIA DE FINANCIAMENTO DO CÓDIGO IMPLÍCITOS
17  NOME-AGENCIA CDATA IMPLIED
18  ANO-DE-OBTENCAO-DO-TITULO CDATA IMPLIED
19  TITULO-DA-DISSERTACAO-TESE CDATA IMPLIED
20  NOME-COMPLETO-DO-ORIENTADOR CDATA IMPLIED
21  TIPO-DOUTORADO CDATA IMPLIED
22  CODIGO-INSTITUICAO-DOUT CDATA IMPLIED
23  NOME-INSTITUICAO-DOUT CDATA IMPLIED
24  CODIGO-INSTITUICAO-OUTRA-DOUT CDATA IMPLIED
25  NOME-INSTITUICAO-OUTRA-DOUT CDATA IMPLIED
26  NOME-ORIENTADOR-DOUT CDATA IMPLIED
27  CDATA NÚMERO-ID-ORIGINADOR #IMPLIED
28  CODIGO-CURSO-CAPES CDATA IMPLIED
29  TITULO-A-DISSERTACAO-ESTE-INGLÊS CDADOS IMPLÍCITOS
30  NOME-CURSO-INGLES CDATA IMPLIED
31  NOME-DO-ORIENTADOR-CO-TUTELA CDATA IMPLIED
32  CODIGO-INSTITUICAO-OUTRA-CO-TUTELA CDATA IMPLIED
33  CODIGO-INSTITUICAO-CO-TUTELA CDATA IMPLIED
34  NOME-DO-ORIENTADOR-SANDUICHE CDATA IMPLIED
35  CODIGO-INSTITUICAO-OUTRA-SANDUICHE CDATA IMPLIED
36  CODIGO-INSTITUICAO-SANDUICHE CDATA IMPLIED
37  NOME-DO-CO-ORIENTADOR CDATA #IMPLIED
38 -->

```

Fonte: elaborada pelos autores.

Essa estrutura apresentada na figura anterior demonstra o elemento “doutorado”, seus atributos e subelementos, presente no arquivo de metadados disponível no Currículo Lattes do pesquisador, no qual sua estrutura (arquivo XML) é composto por elementos (*tags*) iniciados com marcador <elemento> e finalizados com </elemento>, que delimitam o agrupamento dos dados, ou a organização do documento, permitindo aninhamento de outros documentos, mas considerando que para cada arquivo XML só deve haver um marcador raiz.

Por fim, é possível especificar um ou mais atributos para o elemento, mas não pode ocorrer repetição de nome de atributo no mesmo elemento. Além disso, é possível especificar um ou mais subelementos e estes podem ocorrer quantas vezes forem necessárias, por exemplo, autoria em publicação

científica, em que o subelemento é o autor/coautor e pode registrar vários, de acordo com a política do periódico científico. Assim, o elemento “doutorado”, atributos e subelementos (DTD) descreve o fragmento do modelo de dados disponível no arquivo XML do Currículo Lattes.

Considerações finais

A produção e comunicação científica vêm se transformando nas últimas décadas em função do uso de tecnologia da informação e comunicação, que proporciona um canal formal de avaliação dos trabalhos submetidos pelos pesquisadores e permite o acesso aos resultados de pesquisa disponíveis na *web* em formato digital. Essa transformação vem ocorrendo na perspectiva do uso do recurso informacional, que anteriormente era físico e passa para o formato digital. Dessa forma, o que torna o momento ímpar é a informação digital da ciência. Nesse sentido, houve a preocupação em discorrer sobre a coleta de informações sobre artigos científicos registradas no Currículo Lattes para compreender a dinâmica da ciência e buscar formas para contribuir com a melhoria do acesso, por meio do armazenamento das informações de artigos científicos em que a estruturação com metadados e descrição de conteúdos em XML auxiliam nesse processo.

É possível recuperar os dados das comunicações científicas registrados pelos pesquisadores por meio do arquivo XML disponível na Plataforma Lattes. Embora os dados estejam acessíveis, ainda há limitação quanto ao acesso de forma automatizada ou mesmo à possibilidade de intercambiar os dados. A ferramenta Lattes Extrator – que consiste em uma ferramenta de extração dos dados de forma célere – possui restrição quanto à coleta dos dados contidos no Currículo Lattes, visto que há na orientação de uso que a instituição autorizada para utilizar a ferramenta só colete os dados de pessoas que possuem vínculo com a instituição, abstendo-se de acessar os currículos de pessoas externas.

O formato XML utilizado para disponibilizar metadados do Lattes representa as marcações (*tags*) necessárias para que um sistema informatizado consiga processar e identificar cada ação que consta registrada no currículo.

Destaca-se a importância desse sistema de informação na padronização do currículo do pesquisador, bem como para identificar os atores envolvidos no desenvolvimento de pesquisa e na comunicação científica.

Referências

BUNGE, M. *Emergence and convergence: Qualitative Novelty and the Unity of Knowledge*. Canadá: University of Toronto Press, Scholarly Publishing Division, 2014.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. *Sobre a plataforma Lattes*. Brasília, DF, 2024a. Disponível em: <http://memoria.cnpq.br/web/portal-lattes/sobre-a-plataforma>. Acesso em: 17 abr. 2024.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. *CNPq anuncia inclusão do campo licença-maternidade no Currículo Lattes*. Brasília, DF, 2024b. Disponível em: <https://www.gov.br/cnpq/pt-br/assuntos/noticias/cnpq-em-acao/cnpq-anuncia-inclusao-do-campo-licenca-maternidade-no-curriculo-lattes>. Acesso em: 17 abr. 2024.

DAPPERT, A.; ENDERS, M. Digital Preservation: Metadata Standards. *Information Standards Quarterly*, [s. l.], v. 22, n. 2, p. 4-13, 2010. Disponível em: https://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf. Acesso em: 22 jul. 2022.

DIGIAMPIETRI, L. A. *et al.* Análise macro das últimas atualizações dos Currículos Lattes. *Em Questão*, Porto Alegre, v. 20, n. 3, p. 88-113, 2014. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/49173/34060>. Acesso em: 30 out. 2022.

KERN, V. M. Plataformas e-GOV como sistemas sociotecnológicos. In: ROVER, A. J.; GALINDO, F. (ed.). *O Governo Eletrônico e suas múltiplas facetas*. Zaragoza: Prensas Universitarias de Zaragoza, 2010. p. 39-68. Disponível em: <https://zaguan.unizar.es/record/5995/files/BOOK--2011-001.pdf?version=1>. Acesso em: 6 set. 2022.

LANE, J. Let's make science metrics more scientific. *Nature*, [s. l.], v. 464, p. 488-489, Mar. 2010. DOI 10.1038/464488a. Disponível em: <https://www.nature.com/articles/464488a>. Acesso em: 6 out. 2023.

MARCONDES, C. H.; SAYÃO, L. F. Documentos digitais e novas formas de cooperação entre sistemas de informação em C&T. *Ciência da Informação*, Brasília, DF, v. 31, n. 3, p. 42-54, set./dez. 2002. DOI 10.1590/S0100-19652002000300005. Disponível em: <https://www.scielo.br/j/ci/a/NKhjHgVf63bYGmkHJWQkWhB/?format=pdf&lang=pt>. Acesso em: 21 maio 2018.

MEADOWS, A. J. *A comunicação científica*. Tradução de Antonio Agenor Briquet Lemos. Brasília, DF: Briquet de Lemos, 1999.

MENA-CHALCO, J. P. *Definição de bancas de concurso docente*. [S. l.], 9 nov. 2015. Disponível em: <http://scriptlattesblog.blogspot.com/2015/11/definicao-de-bancas-de-concurso-docente.html>. Acesso em: 17 abr. 2024.

MENA-CHALCO, J. P.; CESAR-JUNIOR, R. M. *Uma ferramenta para extração e visualização de conhecimento a partir de Currículos Lattes*. [S. l., 20--]. Disponível em: <http://scriptlattes.sourceforge.net/>. Acesso em: 17 abr. 2024.

MILLER, E. An Introduction to the Resource Description Framework. *Bulletin of the American Society for Information Science*, Washington, v. 4, n. 5, p. 1-10, Oct./Nov. 1998. DOI 10.1002/bult.105. Disponível em: <https://asistdl.onlinelibrary.wiley.com/doi/epdf/10.1002/bult.105>. Acesso em: 21 maio 2018.

MUGNAINI, R.; LEITE, P.; LETA, J. Fontes de informação para análise de internacionalização da produção científica brasileira. *PontodeAcesso*, Salvador, v. 5, n. 3, p. 87-102, ago. 2011. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/5684/4108>. Acesso em: 10 out. 2024

MUGNAINI, R. *et al.* Panorama da produção científica do Brasil além da indexação: uma análise exploratória da comunicação em periódicos. *TransInformação*, Campinas, v. 31, 2019. DOI 10.1590/2318-0889201931e190033. Disponível em: <https://www.scielo.br/j/tinf/a/gWy9yV67t3WHRmWdG8DP7qS/?format=pdf&lang=pt>. Acesso em: 3 abr. 2021.

NATIONAL INFORMATION STANDARDS ORGANIZATION. Digital identifiers roundtable reports on findings and recommendations for NISO. *Information Standards Quarterly*, [s. l.], v. 18, n. 3, p. 1-6, 2006.

- PACHECO, R.; KERN, V. Arquitetura conceitual e resultados da integração de sistemas de informação e gestão da ciência e tecnologia. *DataGramaZero - Revista de Ciência da Informação*, [s. l.], v. 4, n. 2, abr. 2003. Disponível em: <http://afro.culturadigital.br/wp-content/uploads/2017/10/Artigo-28.pdf>. Acesso em: 5 abr. 2021.
- RODRIGUES, W. B.; TOMASSINI-URTI, M. C.; MENA-CHALCO, J. P. Análise da estrutura de redes de sugestão de pareceristas científicos. In: ENCONTRO NACIONAL DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 23., 2023, São Cristóvão/SE. *Anais [...]*. São Cristóvão: UFS, 2023. Disponível em: <https://ancib.org/enancib/index.php/enancib/xxxiiienancib/paper/viewFile/1826/1200>. Acesso em: 17 abr. 2024.
- SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. *Sistema de banco de dados*. 5. ed. Rio de Janeiro: Elsevier, 2006.
- SILVA, F. M.; SMIT, J. W. Organização da informação em sistemas eletrônicos abertos de Informação Científica & Tecnológica: análise da Plataforma Lattes. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 14, n. 1, p. 77-98, jan./abr. 2009. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/23584/19047>. Acesso em: 11 mar. 2021.
- SILVA, F. M. *et al.* Mapeamento da produção científica brasileira sobre acesso aberto: 2001 a 2011. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 17, p. 35-19, 2012. Número especial. DOI 10.5007/1518-2924.2012v17nesp2p19. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17nesp2p19/23566>. Acesso em: 15 mar. 2023.
- SOUSA, J. L.; MARTINS, P. G. M.; RAMALHO, R. A. S. Análise dos padrões XML e RDF para a representação na web sob a perspectiva da Ciência da Informação: um estudo preliminar. *Informação & Tecnologia*, Marília, v. 5, n. 1, p. 65-78, jan./jun. 2018. DOI 10.22478/ufpb.2358-3908.2018v5n1.38458. Disponível em: <https://periodicos.ufpb.br/index.php/itec/article/view/38458/22109>. Acesso em: 6 set. 2019.
- TARGINO, M. das G. Comunicação científica: uma revisão de seus elementos básicos. *Informação & Sociedade: Estudos*, João Pessoa, v. 10, n. 2, p. 1-27, 2000. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/326/248>. Acesso em: 22 nov. 2017.

Sobre os autores

Adilson Luiz Pinto

Graduado em Biblioteconomia e mestre em Ciência da Informação pela Pontifícia Universidade Católica (PUC) Campinas. Mestre em Documentação Audiovisual e doutor em Documentação pela Universidad Carlos III de Madrid. Professor associado III do Departamento de Ciência da Informação da Universidade Federal de Santa Catarina (UFSC): graduação em Biblioteconomia/Arquivologia/Ciência da Informação e do mestrado/doutorado em Ciência da Informação. Temas de interesse: bibliometria, cienciométrica, informetria, *librametrics*, webometria e arquivometria, *e-science*, análise de redes sociais, visualização analítica, visualização de textos, modelagem de processo, visualização temporal, geolocalização de dados e clusterização. <http://lattes.cnpq.br/4767432940301118>

Amanda Ramalho

Graduada em Biblioteconomia e pós-graduada em Docência no Ensino Superior pelas Faculdades Integradas Coração de Jesus (Fainc). Mestre em Ciência da Informação pela Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP). Coordenadora do Projeto Scientific Electronic Library Online (SciELO) Livros, sendo responsável pelo gerenciamento da produção dos *e-books* e pela exportação de metadados para bases indexadoras nacionais e internacionais. <http://lattes.cnpq.br/3021040423162521>

Christian Hauschke

Trabalhou por mais de 20 anos em bibliotecas acadêmicas em Magdeburg, Düsseldorf e Hanover em várias áreas – serviços de descoberta, repositório de acesso aberto e consultoria etc. Desde 2016, ele está no TIB Open Science Lab, onde lidera o Lab Group Open Research Information desde 2020. Além disso, ele é membro de vários comitês e iniciativas na área de informações de

pesquisa – por exemplo, DINI AG Forschungsinformationen und Systeme e Comitê Técnico do International Organisation for Research Information (euroCRIS). <https://orcid.org/0000-0003-2499-7741>

Dalton Lopes Martins

Graduado em Engenharia Elétrica e mestre em Engenharia da Computação pela Universidade Estadual de Campinas (Unicamp). Doutor em Ciências da Informação pela Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP). Professor no curso de Biblioteconomia e do Programa de Pós-Graduação em Ciência da Informação (PGGCinf) da Universidade de Brasília (UnB). Também é docente no Programa de Pós-Graduação em Estudos da Condição Humana da Universidade Federal de São Carlos (UFSCar). Temas de interesse: mapeamento, análise estrutural e dinâmica de redes sociais; objetos e repositórios digitais; acervos digitais e estratégias de interoperabilidade; dados abertos ligados; ciência de dados e aprendizagem de máquina, com ênfase na análise de objetos digitais. <http://lattes.cnpq.br/3774617443225038>

Edna da Silva Angelo

Graduada em Biblioteconomia pela Universidade Federal de Minas Gerais (UFMG). Bibliotecária da Imprensa Nacional da Presidência da República, Brasília, Distrito Federal, Brasil. Master in Business Administration (MBA) em Marketing pelo Centro Universitário Una. Mestre e doutora em Gestão e Organização do Conhecimento pela UFMG. <http://lattes.cnpq.br/2724955848652904>

Ednéia Silva Santos Rocha

Graduada em Biblioteconomia pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp). Mestre em Ciência, Tecnologia e Sociedade pela Universidade Federal de São Carlos (UFSCar). Doutora em Política Científica e Tecnológica pela Universidade Estadual de Campinas (Unicamp). Docente no curso de Biblioteconomia e Ciência da Informação na Faculdade de

Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo (FFCLRP/USP); e Programa de Pós-Graduação em Ciência da Informação da Escola de Comunicações e Artes (ECA/USP). Temas de interesse: indicadores bibliométricos, análise de produção científica, altmetria, acesso aberto, integridade científica, informação jurídica e formação do profissional da informação. <http://lattes.cnpq.br/0671510926276134>

Fabiana Andrade Pereira

Graduada em Biblioteconomia pela Fundação Escola de Sociologia e Política de São Paulo (FESPSP). Especialista em Gestão da Comunicação em Mídias Digitais pelo Serviço Nacional de Aprendizagem Comercial de São Paulo (Senac/SP). Mestre em Ciência da Informação pela Universidade de São Paulo (USP). Experiência em *big data*, algoritmos e dataficação, mídias sociais/digitais e Search Engine Optimization (SEO). <http://lattes.cnpq.br/9513310472160389>

Fernando Modesto

Graduado em Biblioteconomia e Documentação e mestre em Biblioteconomia pela Pontifícia Universidade Católica (PUC) Campinas. Doutor em Ciências da Comunicação pela Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP). Pós-doutorado pela Universidade Carlos III de Madrid, Espanha. Professor no curso de Biblioteconomia da ECA/USP. Temas de interesse: automação de bibliotecas e de serviços de informação (*software* comercial e *software* livre); *linked data*; dados abertos (Open Data); formatos de intercâmbio bibliográfico; metadados descritivos; repositórios digitais; representação descritiva (modelos conceituais, RDA, BIBFRAME, AACR2); mídias sociais, internet das coisas, inteligência artificial aplicadas ao contexto da informação registrada. <http://lattes.cnpq.br/1070631453914536>

Grischa Fraumann

Graduado em Estudos Hispânicos, Sociologia e História pela Universidade de Mannheim e pela Universidade de Deusto, em Bilbao. Mestre em Pesquisa e Inovação no Ensino Superior, do programa Mestrados Conjuntos Erasmus Mundus, pela Universidade do Danúbio, em Krems, pela Universidade de Tampere e pela Universidade Normal de Pequim. Doutorando no Departamento de Comunicação da Universidade de Copenhague. Estágios de pesquisa na Universidade de Leiden e na Universidade de Lund. É pesquisador afiliado do Grupo de Pesquisa em Métricas de Ciência e Tecnologia (CiMetrias), da Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP). Temas de interesse: altmetria, pesquisa em ensino superior e política de pesquisa. <https://orcid.org/0000-0003-0099-6509>.

Henrique Baraldi Spera

Graduado em Biblioteconomia pela Universidade de São Paulo (USP). Mestre em Ciência da Informação pela USP. Doutor em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação (PPGCI) da Escola de Comunicações e Artes (ECA) da USP. Bibliotecário da Fundação Getúlio Vargas (FGV/SP). Membro da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (Ancib). Temas de interesse: produção científica, desenvolvimento científico, métodos quantitativos, bibliometria. <http://lattes.cnpq.br/0291859980691667>

Jean Carlos Borges Brito

Graduado em Sistemas de Informação pela Faculdade Cenecista de Brasília (Faceb). Pós-graduado em Gestão Estratégica pela Universidade Cândido Mendes (UCAM); e em Gerência de Projetos com Ênfase em Sistemas de Informação pela Faculdade Santa Terezinha (FST). Mestre em Gestão do Conhecimento e da Tecnologia da Informação pela Universidade Católica de Brasília (UCB). Doutorando em Ciência da Informação pela Universidade de Brasília (UnB). Coordenador de Tecnologia da Informação e Comunicação na Agência Espacial Brasileira (AEB). Experiências nas

áreas de infraestrutura de Tecnologia da Informação (TI); desenvolvimento/programação de sistemas; telecomunicações; auditoria de TI; capacitação e gestão de projetos; governança de TI; e alinhamento estratégico entre TI e negócios. <http://lattes.cnpq.br/1005167212309269>

Lúcia Maria Sebastiana Verônica Costa Ramos

Graduada em Biblioteconomia e Ciência da Informação pela Fundação Escola de Sociologia e Política de São Paulo. Mestre e doutora em Ciência da Informação pela Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP). Diretora técnica do Serviço de Documentação Odontológica da Faculdade de Odontologia da USP. Coordenadora da BVS Odontologia Brasil. Temas de interesse: bibliotecas especializadas; odontologia; divulgação científica; acesso à informação e redes de informação científica; e gestão dos serviços da biblioteca. <http://lattes.cnpq.br/4420248219310062>

Marcos Luiz Mucheroni

Graduado em Ciência da Computação pela Universidade Federal de São Carlos (UFSCar). Mestre em Engenharia Mecânica e doutor em Engenharia Elétrica pela USP. Professor do curso de Biblioteconomia da Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP). Docente e orientador nos Programas de Pós-Graduação em Ciências da Informação (PPGCI); e Mestrado Profissional em Gestão da Informação, ambos da ECA/USP. Temas de interesse: arquiteturas de computadores, sistemas distribuídos, sistemas operacionais, linguagens de programação, processamento de imagens, *web* semântica, educação à distância, redes sociais, cibercultura e organização do conhecimento. <http://lattes.cnpq.br/7224504768054079>

Marlene Oliveira

Graduada em Biblioteconomia pelo Centro Universitário de Formiga (Minas Gerais). Mestre pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Rio de Janeiro). Doutora pela Universidade de Brasília (UnB).

Docente do Programa de Pós-Graduação em Gestão e Organização do Conhecimento na Escola de Ciência da Informação da Universidade Federal de Minas Gerais (UFMG). <http://lattes.cnpq.br/0146077522102988>

Rogério Mugnaini

Graduado em Estatística. Mestre em Ciência da Informação pela Pontifícia Universidade Católica (PUC) Campinas. Doutor em Ciência da Informação pela Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP). Professor do curso de Biblioteconomia da ECA/USP. Coordenador do Programa de Pós-Graduação em Ciência da Informação da ECA/USP. Coordenador do Grupo de Pesquisa em Métricas da Ciência e Tecnologia (CiMetrias) da ECA/USP. Estágios de pesquisa nas universidades: Universidad Carlos III de Madrid e Leiden University. Temas de interesse: bibliometria; cientometria; avaliação de produção científica nacional; indicadores; fontes de informação; política científica e ciência aberta. <http://lattes.cnpq.br/6156509933307243>

Rubenildo Oliveira da Costa

Graduado em Ciência da Informação e Biblioteconomia e mestre em Ciência da Informação pela Pontifícia Universidade Católica (PUC) Campinas. Bibliotecário no Centro de Energia Nuclear na Agricultura da Universidade de São Paulo (USP). Editor administrativo da Autopsy and Case Reports. Temas de interesse: produção e comunicação científica; bibliometria, com aplicação dos *softwares* BibExcel, VantagePoint e Pajek; metodologia Scientific Electronic Library Online (SciELO) do Portal de Revistas de Odontologia Rev@Odonto. <http://lattes.cnpq.br/0055286650034408>

Solange Maria dos Santos

Graduada em Biblioteconomia e mestre e doutora em Ciência da Informação pela Escola de Comunicações e Artes da Universidade de São Paulo (ECA/USP). Membro do Colegiado de Gestão do programa Scientific Electronic Library Online (SciELO). Coordenadora de produção e publicação da rede

SciELO. Membro do Advisory Board do Directory of Open Access Journals (DOAJ). Temas de interesse: comunicação científica; revistas científicas; acesso aberto; produção científica; indicadores bibliométricos; e *rankings* de universidades. <http://lattes.cnpq.br/1490307259359399>

Svantje Lilienthal

Graduação e mestrado em Library and Information Science e Computer Science pela Humboldt-Universität zu Berlin. Trabalhou em aplicação para a *web* semântica: TripleGeany, que é um aplicativo que auxilia o usuário na geração, manipulação e visualização de triplas RDF. Aplicativo desenvolvido no Laboratório Interdisciplinar de Imagem, Conhecimento e Design da Humboldt-Universität. Atualmente, desenvolve projetos para Wikimedia Deutschland. <https://orcid.org/0000-0003-1537-2862>

Thiago Magela Rodrigues Dias

Graduado em Ciência da Computação pelo Centro Universitário de Formiga (Unifor). Especialização em Produção de Software e em Melhoria do Processo de Software pela Universidade Federal de Lavras (UFAL). Mestre e doutor em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (Cefet/MG). Professor no curso de graduação e de pós-graduação em Modelagem Matemática e Computacional do Cefet/MG. Professor do Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina (UFSC). Temas de interesse: bibliometria; cientometria; extração e integração de dados; análise de redes sociais; análise de redes de colaboração científica; acesso aberto; recuperação e organização da informação; ciência de dados; *data mining* e *web mining*. <http://lattes.cnpq.br/4687858846001290>

Valmir Batista Prestes e Souza

Graduado em Informática pela Universidade Federal de Rondônia (UNIR). Especialista em Banco de Dados pela Universidade Luterana do Brasil (Ulbra). Mestre em Administração pela UNIR. Doutor em Ciência da

Informação pela Universidade de São Paulo (USP). Professor do curso de Biblioteconomia da UNIR. Docente e orientador no Mestrado Profissional em Administração Pública (Profiap/UNIR). Chefe do Departamento Acadêmico de Ciências da Computação da UNIR. Editor da revista eletrônica *RARA* (www.rara.unir.br). Temas de interesse: Bancos de dados e *web* semântica. <http://lattes.cnpq.br/3052341015447331>

Vivian dos Santos Silva

Graduada em Ciência da Computação pela Universidade Federal do Rio de Janeiro (UFRJ). Mestre em Informática pela UFRJ. Doutora pela Universidade de Passau (Alemanha), onde fez parte do Semantic Computing and Natural Language Processing Group. Pesquisadora no Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Experiência como analista de sistemas em empresas dos setores público, educacional, elétrico e bancário. Temas de interesse: processamento de linguagem natural, extração e representação de conhecimento, inferência semântica sobre conhecimento de senso comum, interpretabilidade semântica, grafos de conhecimento, e modelagem conceitual de dados. <http://lattes.cnpq.br/6480849016794725>

Washington Luís R. de Carvalho Segundo

Graduado em Matemática; mestre e doutor em Informática pela Universidade de Brasília (UnB). Estágio de doutorado sanduíche na King's College London. Coordenador técnico da área de tratamento, análise e disseminação da informação científica no Instituto Brasileiro de Informação em Ciência e Tecnologia do Ministério de Ciência, Tecnologia e Inovação (IBICT/MCTI). Líder do Grupo de Pesquisa e Laboratório do Ecossistema da Pesquisa Científica Brasileira (LaEPeCBr). Temas de interesse: métodos formais; repositórios digitais abertos; repositórios de dados científicos; interoperabilidade entre sistemas de informação abertos; ciência aberta e ciência de dados. <http://lattes.cnpq.br/9453481318889500>